Conformations in Solution and in Solid-State Polymorphs: Correlating Experimental and Calculated NMR Chemical Shifts for Tolfenamic Acid

Helen Blade[a*], Charles D. Blundell[b*], Steven P. Brown[c*], Jake Carson[d], Hugh R. W. Dannatt[b], Leslie P. Hughes[a*] and Anjali K. Menakath[c]

[a] Oral Product Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Macclesfield, UK

[b] C4X Discovery, Manchester, M1 3LD, UK

[c] Department of Physics, University of Warwick, Coventry, CV4 7AL, UK

[d] Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

[*] Corresponding authors

E-mail: S.P.Brown@warwick.ac.uk

E-mail: Leslie.Hughes2@astrazeneca.com

Email: Charles.Blundell@c4xdiscovery.com

Email: Helen.Blade@astrazeneca.com

Abstract

A new approach for quantitively assessing putative crystal structures with applications in crystal structure prediction (CSP) is introduced that is based upon experimental solution- and magic-angle spinning (MAS) solid-state NMR data and density functional theory (DFT) calculation. For the specific case of tolfenamic acid (TFA), we consider experimental solution-state NMR for a range of solvents, experimental MAS NMR of polymorphs I and II, and DFT calculations for four polymorphs. The change in NMR chemical shift observed in passing from the solution state to the solid state ($\Delta\delta_{Experimental}$) is calculated as the difference between ^{1}H and ^{13}C experimental solid-state chemical shifts for each polymorphic form ($\delta_{Solid\ expt}$) and the corresponding solution-state NMR chemical shifts ($\delta_{Solution\ expt}$). Separately, we use the gauge-included projector augmented wave (GIPAW) method to calculate the NMR chemical shifts for each form ($\delta_{Solid\ calc}$) and for TFA in solution ($\delta_{Solution\ calc}$), using the dynamic 3D solution conformational ensemble determined from NMR spectroscopy. The calculated change in passing from the solution state to the solid state, $\Delta\delta_{Calculated}$, is then calculated as the difference of $\delta_{Solid\ calc}$ and $\delta_{Solution\ calc}$. Regression analysis for $\Delta\delta_{Calculated}$ against $\Delta\delta_{Experimental}$ followed by a t-test for statistical significance provides a robust quantitative assessment. We show that this assessment clearly identifies the correct polymorph, *i.e.,* when comparing $\Delta\delta_{Experimental}$ based on the experimental MAS NMR chemical shifts of Form I or II with $\Delta\delta_{Calculated}$ based on calculated chemical shifts for polymorphs I, II, III and IV. Complementarity to the established approach of comparing $\delta_{Solid\ expt}$ to $\delta_{Solid\ calc}$ is explored. We further show that our approach is applicable if there are no solid-state crystal structure data. Specifically, $\delta_{Solid\ calc}$ in $\Delta\delta_{Calculated}$ is replaced by the chemical shift for an isolated molecule with a specific conformation. Sampling conformations at specific 15° angle values and comparing them against experimental ^{13}C chemical shift data for Forms I and II identifies matching narrow ranges of conformations, successfully predicting the conformation of tolfenamic acid in each form. This methodology can therefore be used in crystal structure prediction to both reduce the initial conformational search space and also quantitatively assess subsequent putative structures to reliably and unambiguously identify the correct structure.

1. Introduction

The gold standard technique for crystal structure determination is single crystal x-ray diffraction (SXRD). Often though a crystal of suitable size or quality is unavailable and alternative approaches are required. One alternative diffraction-based approach is to try and solve the structure from powder x-ray diffraction (PXRD), but this can often be difficult and typically results in many putative structural models that are consistent with the powder diffractogram.[1-4]

NMR crystallography,[5-14] the combined use of solid-state NMR data and *ab initio* calculation of NMR parameters, is now widely used in academia and increasingly in industry as a complementary tool for crystal structure determination. It is useful for refining and improving the quality of structural models based on both SXRD and PXRD data.[15] The NMR crystallography approach, however, requires a reasonably good starting structural model which is then geometry-optimised before the calculation of NMR parameters. For those cases where such a starting model cannot be generated from diffraction data then alternative approaches may be computationally intensive, *i.e.* crystal structure prediction (CSP),[16-25] with CSP approaches being invaluable for exploring the conformational scape in the search of new polymorphs and assessing the risk of polymorphism.

Current approaches to link CSP to NMR experiment are based on directly comparing calculated (usually using the gauge-including projector augmented wave, GIPAW, method[26]) and experimental solid-state NMR chemical shifts. Specifically, for organic molecules, Emsley and co-workers have established a threshold root-mean-squared error (RMSE) of 0.33 ppm for ^{1}H.[9,27-29] Additionally bringing in information about C-H proximities from ^{1}H-^{13}C two-dimensional correlation spectra has also been presented.[29,30] For form 4 of the drug 4-[4-(2-adamantylcarbamoyl)-5-tert-butyl-pyrazol-1-yl] benzoic acid, Baias *et al.*[31] started from the known molecular formula and, without any structural hypothesis, trial crystal structures were predicted by exploring the lattice energy surface looking for the most stable local minima. Trial crystal structures were generated independently with each of 80 starting molecular conformations in the 32 most commonly observed space groups. By comparing calculated and experimentally measured NMR chemical shifts, a structural model was proposed; this was the first example of a *de novo* NMR crystal structure determination of a previously unknown structure. This approach, whilst robust, is both computationally and time expensive. Paruzzo *et al.*[32] also note the computational cost associated with NMR crystallography, especially for larger or more complex crystals, and have proposed a machine learning-based approach to predict chemical shifts in solids.

Conformational polymorphism, where a different torsion angle exists in a flexible part of a molecule, adds an additional complication as this cannot be easily modelled using CSP approaches. Cruz-Cabeza *et al.*[33] remind us that intermolecular interactions present in organic crystals are generally not sufficient to significantly perturb bond lengths and bond angles, but for those molecules that do exhibit torsional degrees of freedom, various polymorphs can exhibit different molecular conformations and, more importantly, different conformers. CSP methods rely on the conformation being fixed first and then the conformers are packed; with several rotatable bonds an extremely large set of conformers is often generated, with an attendant even greater set of structural models. CSP also becomes increasingly more complex when the number of molecules in the asymmetric unit cell is greater than one.

In this work, we present a novel approach for quantitatively assessing putative crystal structures using a combination of solution- and solid-state magic-angle spinning (MAS) NMR data. Tolfenamic acid (TFA) is used as a proof-of-concept system because it has several conformational polymorphs with similar packing arrangements[34-36] and two of them can be easily prepared in sufficient quantities for solid-state NMR analysis. By introducing solution-state NMR conformational and chemical shift data, our approach has the potential to enhance the efficiency and accuracy of CSP by reducing the number of possible alternative structures.

Our approach is to measure the ^{1}H and ^{13}C experimental solid-state chemical shifts for each polymorphic form ($\delta_{Solid\ expt}$) of TFA and the corresponding solution-state NMR chemical shifts ($\delta_{Solution\ expt}$), taking differences due to solvent into account. The change in chemical shift experimentally observed in passing from the solution state to the solid state ($\Delta\delta_{Experimental}$) is then calculated by difference, as shown in equation 1. In parallel, we also calculate the NMR chemical shifts for each solid form of tolfenamic acid ($\delta_{Solid\ calc}$) and for TFA in solution ($\delta_{Solution\ calc}$), using the solution conformational ensemble determined from NMR spectroscopy, to derive the calculated change in passing from the solution state to the solid state as given by equation 2 ($\Delta\delta_{Calculated}$). Finally, regression analysis of the calculated against the experimentally observed changes in chemical shifts followed by a t-test on the coefficient of determination for significance gives a quantitative assessment that we will show successfully identifies and readily discriminates between the different polymorphs of TFA.

$$\Delta\delta_{Experimental} = \delta_{Solid\ expt} - \delta_{Solution\ expt} \quad \{1\}$$

$$\Delta\delta_{Calculated} = \delta_{Solid\ calc} - \delta_{Solution\ calc} \quad \{2\}$$

2 Methods

2.1 Materials

TFA was used as received from Sigma-Aldrich (Gillingham, UK). It was confirmed as polymorphic Form I by comparing the experimental PXRD pattern with that predicted, using the program Mercury[37] (Cambridge Crystallographic Data Centre), from the crystal structure in the Cambridge Structural Database[38] (CSD accession number: 1960856), see supplementary Figure S1. Form II was prepared following the method of Du *et al.*[39] Briefly, a saturated solution of Form I in ethyl acetate at 70°C was crash cooled to 5°C (using an ice-bath). To ensure that the correct form was prepared, the saturated solution was seeded with 1 mg of Form II material isolated from an earlier (smaller scale) experiment. The solution was filtered rapidly after crystallisation to prevent conversion to Form I. The experimental PXRD pattern was compared with that predicted from the crystal structures for Form II (CSD accession number: 1960855), see supplementary Figure S1.

Since the other polymorphs of TFA could not easily be prepared, structural data only from two further polymorphs were also used in this work, namely Form III (CSD ref: KAXXAI02) and Form IV (CSD refcode: KAXXAI03).[34]

2.2 Experimental measurement of solid-state NMR chemical shifts

All solid-state NMR experiments were carried out at room temperature using a Bruker Avance III spectrometer operating at a ^{1}H Larmor frequency of 500 MHz (11.7 T) using a 4 mm HXY probe in double resonance mode. A ^{1}H 90 pulse duration of 2.5 μs and a recycle delay of 60 s was used in all experiments.

^{13}C cross polarisation (CP) MAS: CP was achieved by using a 90-100% amplitude ramp[40] on ^{1}H. SPINAL-64 ^{1}H decoupling[41] at a nutation frequency of 100 kHz was applied during an acquisition time of 40 ms.

^{1}H-^{13}C refocused INEPT: A refocused–INEPT[42] pulse sequence was used with ^{1}H homonuclear decoupling (eDUMBO-1_{22})[43] to attain high resolution in the indirect dimension and also to extend the coherence lifetimes during the spin-echo periods. The eDumbo-1_{22} cycle duration was 32 μs and pre-pulses of duration 1.2 μs were used. 8 transients were co-added for each of 256 t_1 FIDs using the

States-TPPI method to achieve sign discrimination, corresponding to a total experimental time of 19 hours.

^{13}C and ^{1}H chemical shifts are referenced with respect to tetramethylsilane (TMS) via L-alanine at natural abundance as a secondary reference (1.1 ppm for the CH_3 ^{1}H resonance and 177.8 ppm for the CO ^{13}C resonance) corresponding to adamantane at 1.85 ppm (^{1}H)[44] and 38.5 ppm (^{13}C).[45] Note that equation 6 of Harris *et al.* (IUPAC Recommendations 2008)[46] makes clear that this commonly used solid-state NMR referencing approach, *i.e.* relative to pure TMS, differs by 0.73 ppm for ^{13}C from a solution of 1% TMS in $CDCl_3$.

2.3 Calculation of solid-state NMR chemical shifts

The crystal structures of each form determined by single-crystal X-ray diffraction (Form 1: CSD accession code 1960856; Form 2: CSD accession code 1960855; Form 3: KAXXAI02; Form 4: KAXXAI03) were used as the starting structure for geometry optimisation, in which the positions of all the atoms in the asymmetric unit were relaxed by keeping the unit cell dimensions fixed. Density functional theory (DFT) calculations were performed using CASTEP[47] Academic Release version 16.1. All calculations used the Perdew Burke Ernzerhof (PBE) exchange correlation functional,[48] a plane-wave basis set with ultrasoft pseudopotentials and a plane-wave cut-off energy of 700 eV. Integrals over the Brillouin zone were taken using a Monkhorst–Pack grid of minimum sample spacing $0.1 \times 2\pi$ Å^{-1}. Figure S2 compares structures before and after geometry optimisation. NMR chemical shielding calculations were carried out on the geometry-optimized structures using the GIPAW[26] method to determine the shielding tensor for each nucleus in the crystal structure.

The new approach presented in this paper that additionally uses solution-state NMR chemical shifts and relies on differences between the solid and solution states can directly use these as-calculated shieldings. However, to allow a comparison of our approach to that based on RMSE between experimental and calculated solid-state NMR chemical shifts,[27] we present calculated isotropic chemical shifts ($\delta_{isocalc}$) as determined from the calculated chemical shieldings (σ_{calc}) using equation 3:

$$\delta_{isocalc} = \sigma_{ref} - m\,\sigma_{calc} \qquad \{3\}$$

where σ_{ref} is the reference shielding. There is considerable discussion in the literature about how such referencing is performed. We use here the simplest approach of taking the sum of the experimental chemical shift and the GIPAW calculated absolute isotropic chemical shieldings.[7,49] This is equivalent to setting the gradient, m, in equation 3 to unity, and results for TFA in σ_{ref} values of 169.9 ppm and 30.0 ppm for ^{13}C and ^{1}H, respectively. An alternative approach would be to take average values over a range of compounds.[50]

Alternative crystal structures for calculation of the NMR chemical shifts for Form I and Form II could have been the lower, standard resolution CSD structures that were also available (KAXXAI01, KAXXAI, respectively), which might have significantly affected the value calculated for the chemical shifts. However, the maximum absolute differences in the calculated solid-state chemical shifts between high and standard resolution structures was 0.1 ppm (^{1}H) and 0.2 ppm (^{13}C) for Form I, and ^{1}H 0.1 ppm (^{1}H) and 0.4 ppm (^{13}C) for Form II (see supplementary Table S1); these are well within the typical uncertainty compared to experiment of 1% of the typical chemical shift range for GIPAW calculations (*i.e.*, 0.2 ppm and 2 ppm for ^{1}H and ^{13}C).[7,11] These differences in chemical shifts are smaller than might have been supposed because the high and standard resolution structures are inherently very similar (root mean standard deviation, $RMSD_1$, of 0.0095 and 0.0109 Å for Forms I and II, respectively). This demonstrates that high resolution crystal structures are unnecessary inputs for this step of the approach, and therefore structures with standard resolution (*i.e.*, checkCIF limit on resolution is 0.84 Å for A-alerts) are fine for this purpose, *i.e.* this approach is more generally useful.[51]

GIPAW NMR calculations were also performed for an isolated single molecule,[52-56] whereby a molecule was extracted from the geometry-optimised crystal structure and placed into a periodically repeating unit cell with dimensions equal to the unit cell plus 10 Å in each direction.

Calculated NMR shieldings were visualised and tabulated using MagresView,[57] a visualization tool developed by CCP-NC (the Collaborative Computational Project for NMR Crystallography (www.ccpnc.ac.uk).

2.4 Powder x-ray diffraction

PXRD analysis was performed by mounting the powder on a silicon wafer mount and using a PANalytical CubiX PRO diffractometer (λ = 1.5418 Å). Samples were measured in reflection geometry over the scan range 2° to 40° 2θ with a nominal 25 second exposure per 0.02° increment. The X-rays were generated by a copper long-fine focus tube operated at 45 kV and 40 mA. Each sample was spun at 30 rpm to improve counting statistics.

2.5 Experimental measurement of solution-state dynamic 3D structure and NMR chemical shifts

The dynamic 3D solution structure of TFA was determined by NMR according to the method of Blundell *et al.*[58] for TFA in water in the charged state, giving the mean angles, % populations and degree of librations for each rotatable bond; this conformational behaviour is summarised in supplementary Table S2 and will be reported in detail by Dannatt *et al* (manuscript in preparation).

Chemical shifts of TFA have been determined in both charged and uncharged states and in a variety of solvents (as listed in Tables in the main text and the SI) – these will be reported in detail by Crick *et al.* (manuscript in preparation). All values used are referenced to the prime internal standard of tetramethylsilane (TMS) in a dilute solution (approximately 1 mM, equivalent to 0.014 wt. %) of $CDCl_3$.[46]

2.6 Calculation of solution-state NMR chemical shifts

Solution-state NMR chemical shifts were calculated by combining measured solution population occupancies for each torsion (as given in supplementary Table S2) and calculated chemical shifts for isolated single molecules as follows:

First, a 'base' conformation of TFA with appropriate bond lengths, angles and improper torsions was generated. There are two high resolution X-ray structures of TFA (CSD accession codes 1960856 and 1960855, respectively) that could provide accurate geometric parameters for this purpose. Since these crystal structures of Form I and Form II are slightly different, both were progressed in parallel as suitable starting points for the best 'base values', i.e., giving two very similar albeit slightly different base geometries (see supplementary Figure S4D). Since the aromatic rings are slightly buckled in both these structures due to packing effects (unlike the mean co-planar conformation in solution), the ring hydrogen and carbon atoms were minimally adjusted for each of the Form I and Form II conformations to ensure all atoms in each ring were co-planar. Additionally, especially in Form II, the R1 N8 atom adopts a hybridisation state with a small component of sp^3 character (see supplementary Figure S4D), representing a perturbation from the solution mean co-planar value (*i.e.*, fully sp^2-hybridised); the improper dihedral angles were therefore minimally adjusted to ensure co-planarity as in solution. For each form, torsion angles R1 O1A – R1 C1 – R1 C2 – R1 C7 and R1 C2 – R1 C7 – R1 N8 – R2 C1 were set to their mean solution conformations of 180°. Finally, the R2 H6A* methyl torsion was set to have one hydrogen directed away from the chlorine atom and co-planar with the R2 ring, as observed in the high-resolution Form I crystal structure (compared to free rotation in solution). The resulting two slightly different 'base' conformations, representing best appropriate estimates of the values present in solution, are herein referred to as 'symmetrised high-resolution' conformations of Form I and Form II.

Second, these symmetrised high-resolution conformations were used to create conformations with R1 C7 – R1 N8 – R2 C1 – R2 C6 set at 15° intervals.

Third, each of these conformations was placed into a periodically repeating unit cell with increased dimensions and the chemical shifts were calculated after first performing a DFT geometry optimisation with the torsion angles fixed using the DFT GIPAW approach described above, representing the calculated chemical shift for a given static conformation in the solution state.

Fourth, these calculated values for static conformations were weighted according to the dynamic solution 3D-structure to produce calculated values for the solution-state NMR chemical shifts ($\delta_{Solution\ calc}$). Solution population occupancies at each 15° value were calculated directly from the continuous Gaussian conformational behaviour given in supplementary Table S2 and these occupancies were used to weight the contribution from each 15° conformation to calculate the solution chemical shift value $\delta_{Solution\ calc}$.

Alternative starting points for creating the 'base' conformation could have been the lower (standard) resolution CSD structures for Form I (KAXXAI01) or Form II (KAXXAI). Comparison of these with the high-resolution X-ray structures is shown in supplementary Figure S4. While the positions of the heavy atoms are essentially unchanged, the hydrogen atom positions are improved in the high-resolution structures throughout, with carbon-hydrogen covalent bonds being 10-15% longer, and bond angle and torsion improvements being seen at the methyl (R2 H6A*) and hydroxyl groups. While these differences between forms might be expected to cause significant differences in the calculated chemical shifts, the DFT geometry optimisation step (*i.e.*, step three above) converges the structures (and especially the hydrogens) such that there is very little difference in the calculated DFT GIPAW chemical shifts (maximum absolute difference of 0.4 ppm for ^{13}C and 0.1 ppm for ^{1}H, see supplementary Table S3). This demonstrates that high resolution crystal structures are also unnecessary inputs for this step of the approach, and therefore structures with standard resolution are fine for this purpose, *i.e.* this approach is more generally useful.[51] A similar result would be expected if the structure of TFA from any other polymorph or co-crystal in the same charge-state were used. For the purposes of this work though, the high-resolution structures have been used for the input because they were available.

2.7 Software

Mercury[37] was used to visualise and compare crystal structures. PyMol[51] was used to create figures. Statistical analysis was undertaken using the R statistical computing platform.[59]

3 Results and discussion

The 2D molecular structure of TFA, its atom numbering and the torsion of interest for this work are shown in Figure 1. In the following discussion, the two aromatic rings are labelled R1 and R2 as shown.

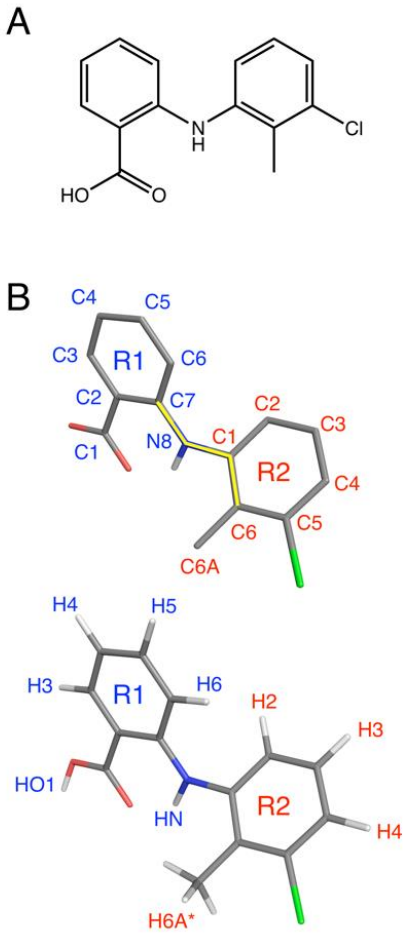


*Figure 1. Tolfenamic acid A) 2D molecular structure, B) atom numbering (top heavy atoms, bottom hydrogens) and torsion of interest (yellow bonds; R1 C7 – R1 N8 – R2 C1 – R2 C6) shown on an example conformation. H6A is labelled with an * to indicate the three hydrogen atoms which each have the same label because they are degenerate both in solution and in the solid state (due to fast rotation).*

3.1 Method Step 1: Conformational behaviour in solution & measurement of solution-state NMR chemical shifts for TFA

The first stage of the approach was to analyse TFA's conformational behaviour in solution and to measure experimentally suitable solution-state chemical shifts for each atom ($\delta_{Solution\ expt}$).

The conformational behaviour for each rotatable bond in TFA in water in the charged state has been determined by NMR according to the method of Blundell *et al.*,[58] giving the mean angle, % population and degree of libration for each possible torsion; this is summarised in supplementary Table S2 (these will be reported in detail by Dannatt *et al*, manuscript in preparation). The structures showing the three conformationally pertinent torsion angles for TFA are also shown in supplementary Figure S3.

This analysis shows that only one torsion angle contributes to the variation in conformation of the molecule and is therefore the only torsion which would result in a perturbation of the solution-state NMR chemical shifts. This torsion is described by the bonds coloured yellow in Figure 1B (top) and in this work is labelled as R1 C7 – R1 N8 – R2 C1 – R2 C6. The dynamic solution 3D-structure shows that this torsion librates about four modes in two symmetrical pairs, with mean angles of *ca*. +/– 135° and +/– 70°. The two pairs of modes are not populated equally (40.7% and 9.3% respectively) and each librate by *ca.* 15°. Figure 2 shows a schematic of the conformational behaviour for this torsion and

Figure 3 shows the relative populations of each mode. None of the other torsions contribute to the overall conformation of the molecule and therefore do not result in chemical shift perturbations. For example, R1 C2 – R1 C7 – R1 N8 – R2 C1 is fixed at 180° (see supplementary Table S2). Comparison of the conformationally-sensitive $^{3}J_{CH}$ coupling constants from R1 HN to R1 C2, R1 C6, R2 C2 and R2 C6 in a variety of solvents and in both neutral and charged states (data not shown – these will be reported in detail by Crick *et al.*, manuscript in preparation) indicates that TFA's conformational behaviour is largely insensitive to both solvent and charge state, *i.e.*, the behaviour determined in the charged state in water is representative of that in other solvents and when neutral.

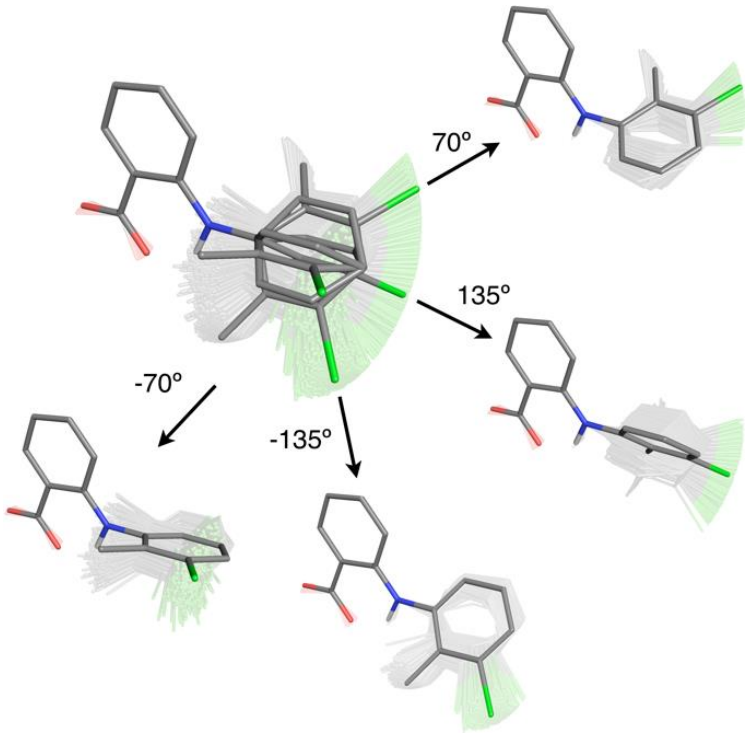


Figure 2. Dynamic solution 3D structure of TFA showing the component four modes for torsion R1 C7 – R1 N8 – R2 C1 – R2 C6 in two symmetrical pairs. The extent of libration is illustrated by the shadows.

Whilst the dynamic solution-state structure is independent of solvent and charge state, nevertheless chemical shifts are solvent and charge-state dependent and therefore an important consideration in measuring them was the choice of an appropriate deuterated solvent that maintained the TFA in the same charge state as present in the crystal forms. Any interaction between the solvent and TFA would perturb the chemical shift and so a solvent system that was barely and non-specifically interacting with TFA was required. For this system, deuterated chloroform was therefore considered to be a suitable starting point for the values for $\delta_{Solution\ expt}$; deutero-benzene and toluene were not initially considered as being suitable due to concerns over specific solvent aromatic ring current effects.

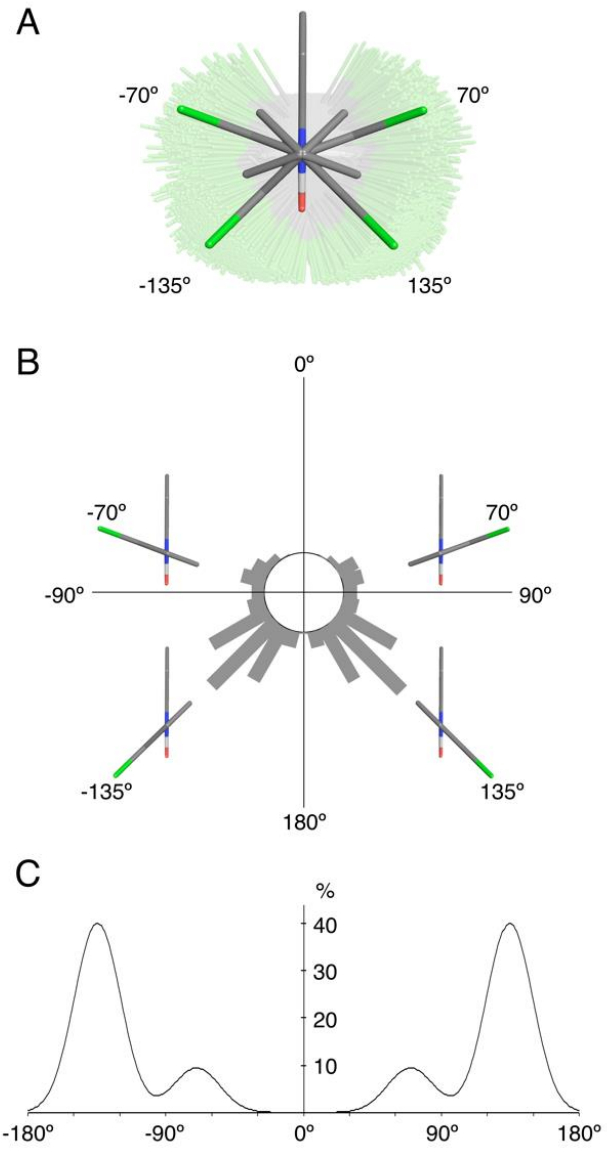


Figure 3. Representations of the dynamic solution 3D structure of TFA which is principally determined by torsion R1 C7 – R1 N8 – R2 C1 – R2 C6. A) 3D structure viewed along bond R2 C1 – R1 N8 with means for each mode emphasised and libration in shadow. B) Circular histogram (15° bins, centred around 0°) representing the population of the conformational distribution; the four modes are shown as insets as 3D structures. C) Continuous graph representing the variation of population with torsion angles of the conformational distribution (maxima correspond to each of the four means).

3.2 Method Step 2: Calculation of solution-state NMR chemical shifts for TFA

The second stage in the approach was to calculate the solution-state chemical shifts for TFA ($\delta_{Solution\ calc}$), using our experimentally derived knowledge of the solution dynamic 3D structure. To perform this calculation, an input 'base' geometry as would be expected to be present in solution was needed (*i.e.* appropriate bond lengths, bond angles, improper dihedrals and with planar rings). The high-resolution X-ray structures of each of Form I and Form II (CSD accession codes 1960856 and 1960855, respectively) provided two suitable starting points for this that differed slightly in their geometries (see supplementary Figure S4). The TFA conformation from each structure was processed to remove geometry distortions caused by packing that would not be present in solution, generating two alternative 'symmetrised high-resolution conformations' for the base geometry (see Experimental). It is noted that the approach does not rely on having high-resolution starting geometries, and that

geometries taken from standard resolution crystal structures and DFT-generated conformations are sufficient (see Experimental).

Using the DFT GIPAW approach, the chemical shifts were calculated for the principal torsion at every 15 degrees for the symmetrised high-resolution conformations from Form 1 and 2 (supplementary Tables S4A and S4B) as gas-phase isolated molecules, *i.e.* for each symmetrised high-resolution conformation, its torsions were set to the relevant values and the molecule was placed into a periodically repeating unit cell with increased dimensions (unit cell dimensions + 10 Å), so as to remove intermolecular interactions. This can be considered to be the calculated chemical shift for a given static conformation in the solution state. As the differences between calculations using Form I and Form II symmetrised high-resolution conformations are small (supplementary Table S4C), the mean calculated chemical shift was used for each static conformation (supplementary Table S4D). As shown in Figure 4 and supplementary Figure S5, ^{1}H and ^{13}C chemical shifts clearly change with torsion angle; moreover, some nuclei are substantially more affected than others (see supplementary Table S5). The variation in chemical shift with torsion angle is different for each nucleus, both in terms of absolute change and profile of behaviour (see Figure 4).

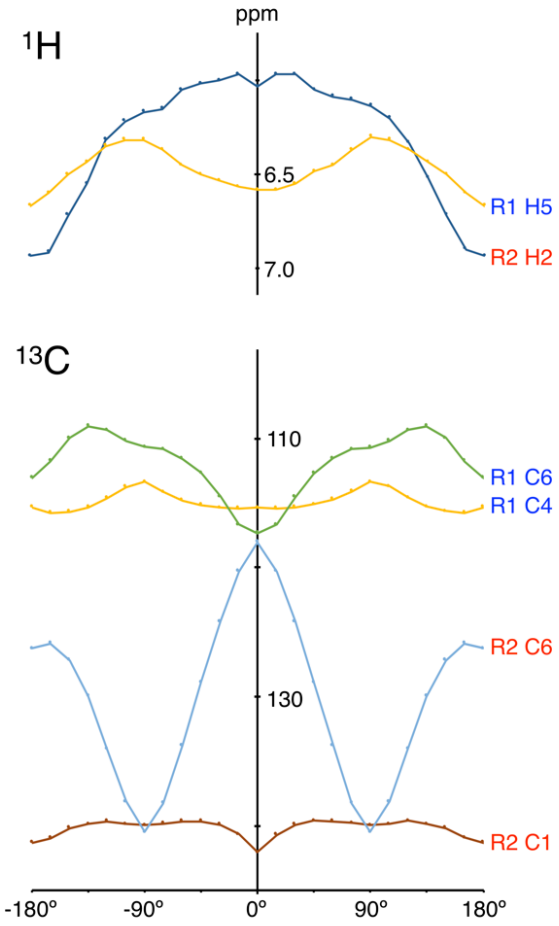


Figure 4. The variation of DFT GIPAW calculated chemical shifts with the R1 C7 – R1 N8 – R2 C1 – R2 C6 torsion angle for selected nuclei in TFA (the variation for all nuclei is shown in supplementary Figure S5). Chemical shifts were calculated for every 15° for the symmetrised high-resolution conformations of both Form I and II as gas-phase isolated molecules and the mean value taken to produce the graph points (see supplementary Table S4D).

Next, the calculated chemical shift values for static conformations were weighted according to the dynamic solution 3D-structure to produce calculated values for the solution chemical shifts ($\delta_{Solution}$

$_{calc}$), as shown in supplementary Table S6. The Gaussian distribution of the population occupancy is shown above supplementary Table S6 as a line graph, with equivalent populations for each 15° histogram bin. The calculated solution value for each nucleus was calculated by weighting the chemical shift with the bin population occupancy. (Note that the discontinuity in calculated chemical shifts at 0°, which is caused by a steric clash, does not contribute to the value of $\delta_{Solution\ calc}$ as this bin is not populated in solution; see, *e.g.,* ^{1}H chemical shifts in Table S4D for R1, H6 and R2, H6A*).

The experimentally measured solution ($\delta_{Solution\ expt}$) and calculated solution ($\delta_{Solution\ calc}$) chemical shifts are shown in the second and fifth columns of Table 1, respectively.

Table 1. Experimentally measured and calculated DFT GIPAW NMR chemical shifts for TFA in solution and for the solid-state forms.

Nucleus		Experimentally measured δ (ppm)			Calculated δ (ppm)									
		$\delta_{Solution\ expt}$	$\delta_{Solid\ expt}$		$\delta_{Solution\ calc}$	$\delta_{Solid\ calc}$								
		($CDCl_3$)	Form I	Form II		Form I	Form II	Form III			Form IV			
		$^1H \pm 0.001$	$^1H \pm 0.2$	$^1H \pm 0.2$										
		$^{13}C \pm 0.02$	$^{13}C \pm 0.1$	$^{13}C \pm 0.1$				*1*[a]	*2*[b]	Mean	*1*[c]	*2*[d]	*3*[e]	Mean
R1	HO1	-[f]	12.6	12.8	4.8	14.0	14.4	*14.0*	*13.4*	13.7	*13.4*	*13.4*	*13.6*	13.5
	H3	8.043	8.4	7.0	7.5	8.0	6.9	*8.0*	*8.0*	8.0	*7.5*	*7.8*	*7.9*	7.7
	H4	6.751	6.6	6.3	5.8	6.6	6.3	*6.4*	*6.3*	6.4	*6.1*	*6.2*	*6.2*	6.1
	H5	7.330	6.1	6.3	6.4	5.9	6.4	*4.9*	*4.6*	4.8	*4.7*	*4.0*	*4.2*	4.3
	H6	6.773	5.6	6.3	6.4	5.2	6.3	*6.8*	*6.8*	6.8	*5.9*	*6.3*	*6.1*	6.1
	NH	9.188	9.0	9.2	10.5	9.7	10.1	*10.5*	*10.9*	10.7	*9.9*	*9.8*	*9.6*	9.8
R2	H2	7.242	6.6	6.2	6.4	6.6	6.2	*6.8*	*6.3*	6.6	*5.9*	*6.9*	*6.6*	6.5
	H3	7.148	5.6	6.1	6.3	5.4	6.0	*5.2*	*6.1*	5.6	*5.8*	*4.7*	*3.9*	4.8
	H4	7.246	6.6	6.3	6.2	6.6	6.2	*6.4*	*6.4*	6.4	*6.0*	*6.7*	*6.1*	6.3
	H6A*	2.335	1.2	1.9	1.4[g]	0.7[g]	1.5[g]	*1.8*[g]	*1.8*[g]	1.8[g]	*0.9*[g]	*1.1*[g]	*0.9*[g]	1.0[g]
R1	C1	172.75	175.2	175.7	170.3	176.3	176.5	*176.0*	*175.8*	175.9	*175.6*	*174.9*	*175.7*	175.4
	C2	109.87	109.2	108.5	106.3	107.8	106.6	*110.1*	*108.8*	109.5	*108.4*	*107.2*	*110.5*	108.7
	C3	132.48	134.9	132.9	132.2	135.8	133.3	*134.5*	*133.7*	134.1	*133.0*	*133.1*	*133.4*	133.2
	C4	117.00	116.5	117.9	115.0	115.6	117.7	*116.7*	*115.6*	116.2	*115.4*	*116.7*	*116.6*	116.2
	C5	135.34	136.9	136.2	134.3	137.5	136.9	*137.0*	*138.2*	137.6	*136.8*	*136.8*	*135.5*	136.4
	C6	113.83	112.0	112.0	109.9	110.5	110.8	*112.2*	*108.7*	110.4	*110.2*	*112.5*	*112.4*	111.7
	C7	149.40	150.4	147.3	147.8	148.8	145.0	*149.3*	*148.6*	148.9	*149.2*	*149.2*	*148.8*	149.1
R2	C1	140.04	138.9	139.1	139.9	138.7	138.1	*140.3*	*139.6*	140.0	*139.4*	*140.3*	*140.7*	140.1
	C2	123.64	127.8	116.8	119.8	126.8	115.2	*122.1*	*123.0*	122.6	*125.7*	*125.7*	*122.7*	124.7
	C3	126.88	127.0	126.0	124.9	127.4	126.3	*127.7*	*128.1*	127.9	*126.3*	*127.5*	*127.3*	127.0
	C4	126.26	129.4	123.3	124.5	129.8	123.5	*126.3*	*126.4*	126.4	*125.6*	*126.0*	*125.9*	125.8
	C5	135.63	138.9	139.1	142.7	140.9	141.5	*142.9*	*140.7*	141.8	*143.4*	*142.9*	*141.0*	142.4
	C6	132.20	136.8	126.1	131.7	138.2	126.4	*131.6*	*132.6*	132.1	*135.7*	*134.3*	*133.7*	134.6
	C6A	15.20	15.4	14.9	9.1	12.2	12.5	*12.7*	*14.3*	13.5	*12.7*	*12.8*	*12.0*	12.5

[a-e]Calculated chemical shift for each conformation from the DFT geometry optimised crystal structure (italics); R1 C7 – R1 N8 – R2 C1 – R2 C6 torsion angles are 126.8°, −138.4°, −115.8°, −126.0° and −134.2° in the diffraction structures, respectively.

[f]Not observable due to chemical exchange with residual water in the solvent.

[g]Mean of chemical shifts for all three hydrogen nuclei for comparison with experimental data.

3.3 Method Step 3: Measurement of solid-state NMR chemical shifts

The third stage of the approach was to measure experimentally the solid-state chemical shifts for Forms I and II ($\delta_{Solid\ expt,}$). The ^{13}C chemical shifts were determined directly by peak-peaking the ^{1}H-^{13}C CP MAS NMR spectrum (see Figure 5A for Form I, and supplementary Figure S6D for Form II) and the ^{1}H chemical shifts were determined from a ^{1}H-^{13}C refocused INEPT MAS NMR spectrum which shows which ^{1}H atom is directly bonded to which ^{13}C atom (see Figure 5B for Form I, and supplementary Figure S7F for Form II). The ^{1}H and ^{13}C solid-state chemical shifts for Forms I and II are given in Table 1 columns 3 and 4 (see supplementary Figure S6 for one-pulse ^{1}H and supplementary Figure S7 for additional two-dimensional ^{1}H-^{1}H double-quantum (DQ) – single-quantum (SQ) and ^{14}N-^{1}H HMQC MAS NMR spectra).

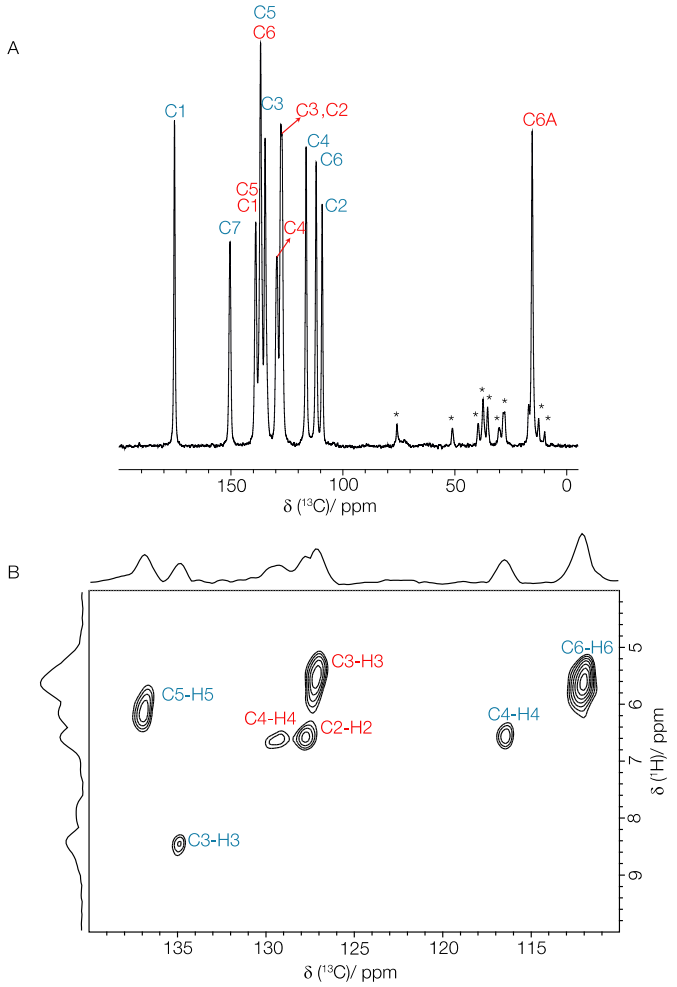


Figure 5. ^{1}H (500 MHz) – ^{13}C MAS (12.5 kHz) NMR spectra of Form I of TFA. A) A CP MAS spectrum recorded with a CP contact time of 1 ms. 1000 transients were co-added with a recycle delay of 60 s. Spinning sidebands are labelled with an asterix. B) An expanded view of the aromatic region of a 2D ^{1}H – ^{13}C refocused INEPT spectrum, with skyline projections, recorded with a spin-echo (τ-π-τ) duration of τ = 1.44 ms and a total experimental time of 19 hours. ^{1}H and ^{13}C chemical shift assignments within each ring are indicated (R1 blue, R2 red). The base contour level is at 37% of the maximum peak height.

3.4 Method Step 4: Calculation of solid-state NMR chemical shifts for TFA

The fourth stage in the approach was to calculate the solid-state NMR chemical shifts ($\delta_{Solid\ calc}$) for each of the four polymorphs using the GIPAW method detailed above. The output of such a GIPAW calculation are chemical shieldings; these are converted into chemical shifts by a referencing procedure. Our new approach takes the difference between calculated values for the solid and solution states, thus negating the need for this referencing step. However, to allow comparison of our method to the established approach of comparing experimental and calculated solid-state NMR chemical shifts, Table 1 shows the chemical shifts calculated for each crystal form as a whole crystal ($\delta_{Solid\ calc}$) using the referencing procedure described in the methods section (see discussion of equation 3). For Form I and II, the high-resolution structures were used because they were available (CSD accession codes 1960856 and 1960855, respectively); the standard resolution crystal structures (KAXXAI01, KAXXAI) would have been equally suitable (see Experimental). Form III and Form IV have multiple conformations in the asymmetric unit and therefore the chemical shifts of each conformation and their mean are presented in Table 1.

Now having the experimentally measured and the calculated chemical shifts for TFA in both solution and the solid state, we are able to quantitatively compare the experimentally measured change (equation 1, $\Delta\delta_{Experimental} = \delta_{Solid\ expt} - \delta_{Solution\ expt}$) with the calculated changes in chemical shifts (equation 2, $\Delta\delta_{Calculated} = \delta_{Solid\ calc} - \delta_{Solution\ calc}$). These changes are given in Table 2. Note that for our simple referencing approach, the $\Delta\delta_{Calculated}$ values are exactly equivalent to the values obtained from the differences in the as-calculated chemical shieldings.

Table 2. Comparison of the experimentally measured ($\Delta\delta_{Experimental}$) and DFT GIPAW calculated ($\Delta\delta_{Calculated}$) changes in NMR chemical shifts (see Table 1) for TFA in passing from the solution state to the solid state.

Nucleus		Experimentally measured change (ppm) $\Delta\delta_{Experimental} = \delta_{Solid\ expt} - \delta_{Solution\ expt}$		Calculated change (ppm) $\Delta\delta_{Calculated} = \delta_{Solid\ calc} - \delta_{Solution\ calc}$								
		Form I	Form II	Form I	Form II	Form III			Form IV			
		$^{1}H \pm 0.2$ $^{13}C \pm 0.1$	$^{1}H \pm 0.2$ $^{13}C \pm 0.1$			1[a]	2[b]	mean	1[c]	2[d]	3[e]	mean
R1	HO1	-[f]	-[f]	9.2	9.6	*9.2*	*8.6*	8.9	*8.6*	*8.6*	*8.8*	8.7
	H3	0.4	-1.0	0.5	-0.6	*0.6*	*0.5*	0.5	*0.0*	*0.3*	*0.4*	0.2
	H4	-0.2	-0.5	0.8	0.5	*0.6*	*0.5*	0.5	*0.3*	*0.4*	*0.4*	0.3
	H5	-1.2	-1.0	-0.5	-0.1	*-1.5*	*-1.8*	-1.6	*-1.7*	*-2.4*	*-2.2*	-2.1
	H6	-1.2	-0.5	-1.2	-0.1	*0.4*	*0.4*	0.4	*-0.5*	*-0.1*	*-0.3*	-0.3
	HN	-0.2	0.0	-0.8	-0.4	*0.0*	*0.4*	0.2	*-0.6*	*-0.7*	*-0.9*	-0.7
R2	H2	-0.6	-1.0	0.1	-0.2	*0.4*	*-0.1*	0.1	*-0.5*	*0.5*	*0.2*	0.0
	H3	-1.5	-1.0	-0.9	-0.3	*-1.1*	*-0.2*	-0.7	*-0.5*	*-1.6*	*-2.4*	-1.5
	H4	-0.6	-0.9	0.3	-0.1	*0.1*	*0.1*	0.1	*-0.3*	*0.5*	*-0.1*	0.0
	H6A*	-1.1	-0.4	-0.7[g]	0.1[g]	*0.4*[g]	*0.4*[g]	0.4[g]	*-0.5*[g]	*-0.3*[g]	*-0.5*[g]	-0.5[g]
R1	C1	2.5	3.0	6.0	6.2	*5.7*	*5.5*	5.6	*5.3*	*4.6*	*5.4*	5.1
	C2	-0.7	-1.4	1.5	0.3	*3.8*	*2.5*	3.2	*2.1*	*0.9*	*4.2*	2.4
	C3	2.4	0.4	3.6	1.1	*2.3*	*1.5*	1.9	*0.8*	*0.9*	*1.2*	0.9
	C4	-0.5	0.9	0.7	2.7	*1.8*	*0.7*	1.2	*0.4*	*1.8*	*1.6*	1.3
	C5	1.6	0.9	3.2	2.6	*2.7*	*3.9*	3.3	*2.5*	*2.5*	*1.2*	2.1
	C6	-1.8	-1.8	0.6	0.9	*2.3*	*-1.2*	0.5	*0.3*	*2.6*	*2.5*	1.8
	C7	1.0	-2.1	0.9	-2.8	*1.4*	*0.8*	1.1	*1.4*	*1.4*	*1.0*	1.2
R2	C1	-1.1	-0.9	-1.2	-1.8	*0.4*	*-0.3*	0.0	*-0.5*	*0.4*	*0.8*	0.2
	C2	4.2	-6.8	7.0	-4.5	*2.3*	*3.3*	2.8	*6.0*	*5.9*	*2.9*	4.9
	C3	0.1	-0.9	2.5	1.4	*2.8*	*3.2*	3.0	*1.4*	*2.6*	*2.4*	2.1
	C4	3.1	-3.0	5.3	-1.0	*1.8*	*2.0*	1.9	*1.1*	*1.5*	*1.4*	1.3
	C5	3.3	3.5	-1.9	-1.2	*0.2*	*-2.0*	-0.9	*0.7*	*0.2*	*-1.8*	-0.3
	C6	4.6	-6.1	6.5	-5.3	*-0.1*	*0.9*	0.4	*4.0*	*2.6*	*2.0*	2.9
	C6A	0.2	-0.3	3.1	3.4	*3.6*	*5.2*	4.4	*3.6*	*3.7*	*2.9*	3.4

[a-e] Calculated change in chemical shift for each conformation in the full crystal ($\delta_{Solid\ calc}$); the R1 C7 – R1 N8 – R2 C1 – R2 C6 torsion angles are 126.8°, –138.4°, –115.8°, –126.0° and –134.2° in the diffraction structures, respectively. [f] Not observable due to chemical exchange with residual water in the solvent in the solution state.
[g] Mean of chemical shifts for all three hydrogen nuclei for comparison with experimental data.

3.5 Method Step 5: Linear regression analysis and t-test to identify the correct form

Finally, we established whether comparison of the values of $\Delta\delta_{Experimental}$ and $\Delta\delta_{Calculated}$ from Table 2 could discriminate the different crystal forms. Figure 6 shows simple linear regression of $\Delta\delta_{Calculated}$ for Forms I, II, III and IV against $\Delta\delta_{Experimental}$ for Form I for both ^{1}H and ^{13}C. The chemical shift difference value for R1 HN has been omitted from the ^{1}H graphs because it was an outlier, presumably due to the solution chemical shift having been affected by solvent exchange. Similarly, the chemical shift difference value for R2 C5 has been omitted from the ^{13}C graphs because it was also an outlier, presumably due to the attached chlorine atom affecting the reliability of the DFT calculation. For each graph, the data has been fitted to a simple linear y = mx + c equation and the coefficient of determination value (r^2) is shown. For Form I, the ^{1}H and ^{13}C experimentally measured change in chemical shift on passing from solution to solid state agrees well with the calculated change (r^2 values of 0.74 and 0.83, respectively). In contrast, the correlation is poor for both ^{1}H and ^{13}C for each of Forms II, III and IV, *e.g.* r^2 value of 0.01 for ^{1}H in Form II. That is, simple comparison of r^2 values alone apparently immediately identifies the correct form.

To establish whether the correlation between $\Delta\delta_{Experimental}$ and $\Delta\delta_{Calculated}$ is indeed able to discriminate between different crystal forms, we perform a t-test with the null hypothesis that m = 0 (no correlation), and the alternative hypothesis that m > 0 (positive correlation). Table 3 shows the p-values obtained for $\Delta\delta_{Experimental}$ for ^{1}H and ^{13}C (Forms I and II) vs $\Delta\delta_{Calculated}$ for Forms I, II, III, and IV. At a significance level of 0.05 we always correctly reject the null hypothesis whenever we have the correct forms. There are some small but significant correlations for conformations from Forms III and IV, especially for ^{1}H, but their p-values are substantially higher than that of the correct Form I. Moreover, considered together with the ^{13}C p-values, the discrimination is unambiguous. Our approach has correctly, quantifiably and unambiguously identified Form I.

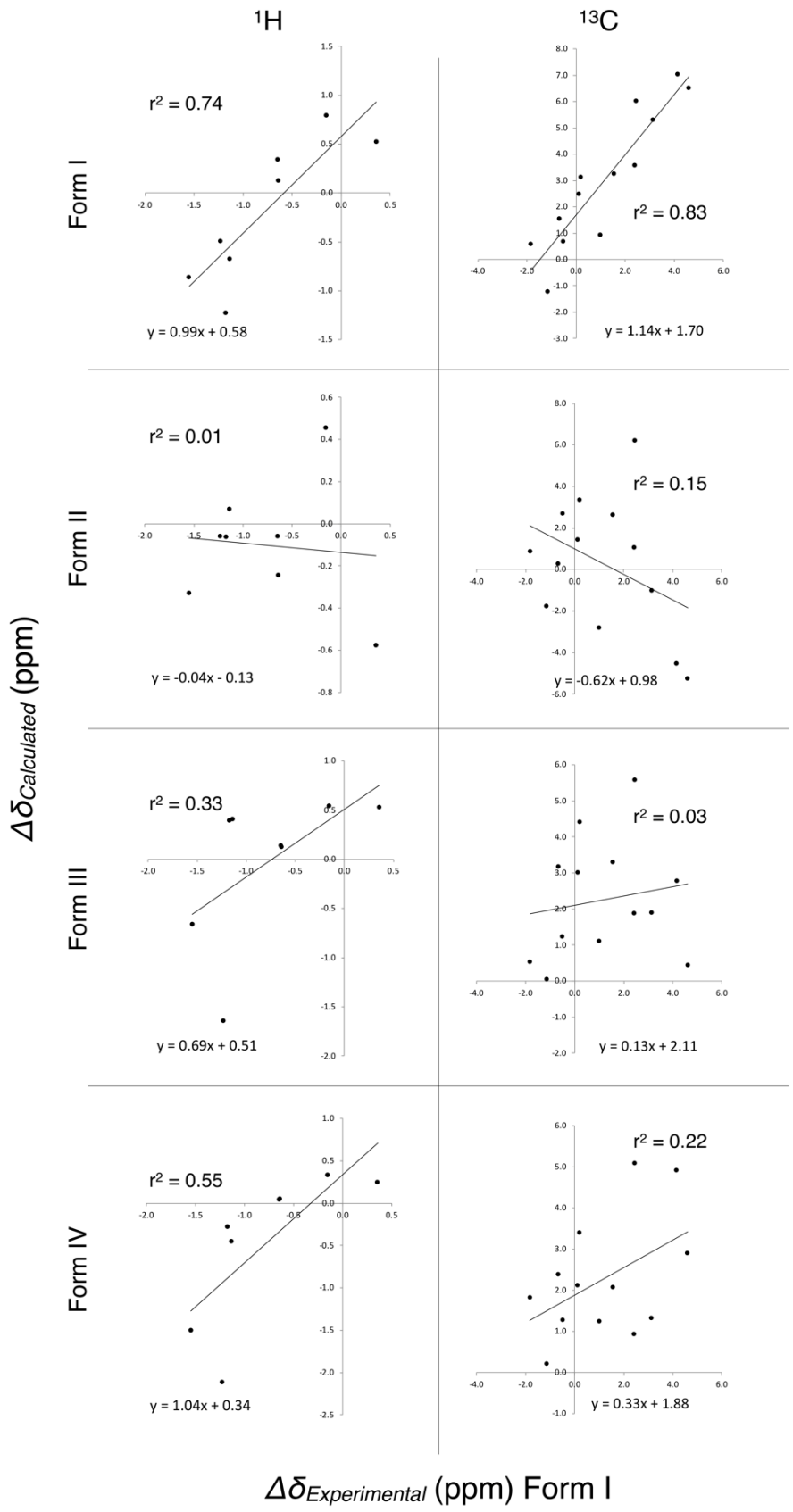


Figure 6. Graph of $\Delta\delta_{Calculated}$ for Forms I, II, III and IV against $\Delta\delta_{Experimental}$ for Form I for both ^{1}H and ^{13}C showing that the approach clearly discriminates Form I ($r^2 = 0.74$ and 0.83 respectively) from the other forms. See Tables 2 and 3 for data.

Table 3. Linear regression analysis parameters and p-values on passing from solution to solid state for combinations of calculated (Form I, II, III and IV) and experimentally measured (Form I and II) changes in chemical shift.

		$\Delta\delta_{Experimental}$ for:							
		Form I				Form II			
$\Delta\delta_{Calculated}$ for:		r^2	m	c	p-value	r^2	m	c	p-value
^{1}H									
Form I		**0.74**[a]	**0.99**	**0.58**	**0.003**[b]	0.03	-0.43	-0.53	0.661
Form II		0.01	-0.04	-0.13	0.586	**0.54**	**0.75**	**0.50**	**0.019**
Form III	mean	0.33	0.69	0.51	0.069	0.27	1.35	1.06	0.091
	1[c]	*0.42*	*0.84*	*0.62*	***0.040***	*0.26*	*1.41*	*1.12*	*0.097*
	2[d]	*0.20*	*0.54*	*0.39*	*0.134*	*0.25*	*1.28*	*1.01*	*0.104*
Form IV	mean	0.55	1.04	0.34	**0.017**	0.11	0.99	0.34	0.211
	1[e]	*0.41*	*0.60*	*-0.01*	***0.042***	*0.11*	*0.65*	*0.05*	*0.213*
	2[f]	*0.46*	*1.15*	*0.55*	***0.033***	*0.08*	*1.04*	*0.49*	*0.247*
	3[g]	*0.60*	*1.38*	*0.49*	***0.012***	*0.12*	*1.29*	*0.47*	*0.204*
^{13}C									
Form I		**0.83**	**1.14**	**1.70**	**<0.001**	0.15	−0.36	2.55	0.902
Form II		0.15	−0.62	0.98	0.905	**0.83**	**1.09**	**1.78**	**<0.001**
Form III	mean	0.03	0.13	2.11	0.301	0.19	0.26	2.63	0.069
	1[c]	*0.01*	*−0.07*	*2.46*	*0.630*	*0.32*	*0.31*	*2.80*	***0.021***
	2[d]	*0.11*	*0.33*	*1.76*	*0.133*	*0.08*	*0.21*	*2.45*	*0.171*
Form IV	mean	0.22	0.33	1.88	0.053	0.03	−0.09	2.16	0.700
	1[e]	*0.44*	*0.65*	*1.40*	***0.007***	*0.08*	*−0.21*	*1.88*	*0.830*
	2[f]	*0.19*	*0.33*	*2.01*	*0.069*	*0.05*	*−0.13*	*2.22*	*0.773*
	3[g]	*0.00*	*0.02*	*2.24*	*0.452*	*0.03*	*0.08*	*2.38*	*0.288*

[a]Values in bold indicate the fit parameters for the form corresponding to the measured experimental data, *i.e.* the ones the approach should identify (see Figures 6, 7, S8 and S9).

[b]p-values are for the null hypothesis that m = 0, and the alternative hypothesis that m > 0. Values in bold and underlined indicate a rejection of the null hypothesis at a significance level of 0.050, suggesting a positive correlation. The lower bound of the one-sided 95% confidence intervals for the correlation between $\Delta\delta_{Experimental}$ and $\Delta\delta_{Calculated}$ are given in supplementary Table S7.

[c-g]R1 C7 – R1 N8 – R2 C1 – R2 C6 torsion angles are 126.8°, −138.4°, −115.8°, −126.0° and −134.2° in the diffraction structures, respectively.

Figure 7 shows $\Delta\delta_{Calculated}$ for Forms I, II, III and IV against $\Delta\delta_{Experimental}$ for Form II for both ^{1}H and ^{13}C (again with R1 HN and R2 C5 omitted). It is again immediately obvious from the r^2 values that Form II is the correct form, even though Forms III and IV have similar conformations (r^2 values of 0.54 and 0.83, respectively) and the p-values (Table 3) clearly quantify this as being statistically significant. Our approach has correctly, quantifiably and unambiguously also identified Form II.

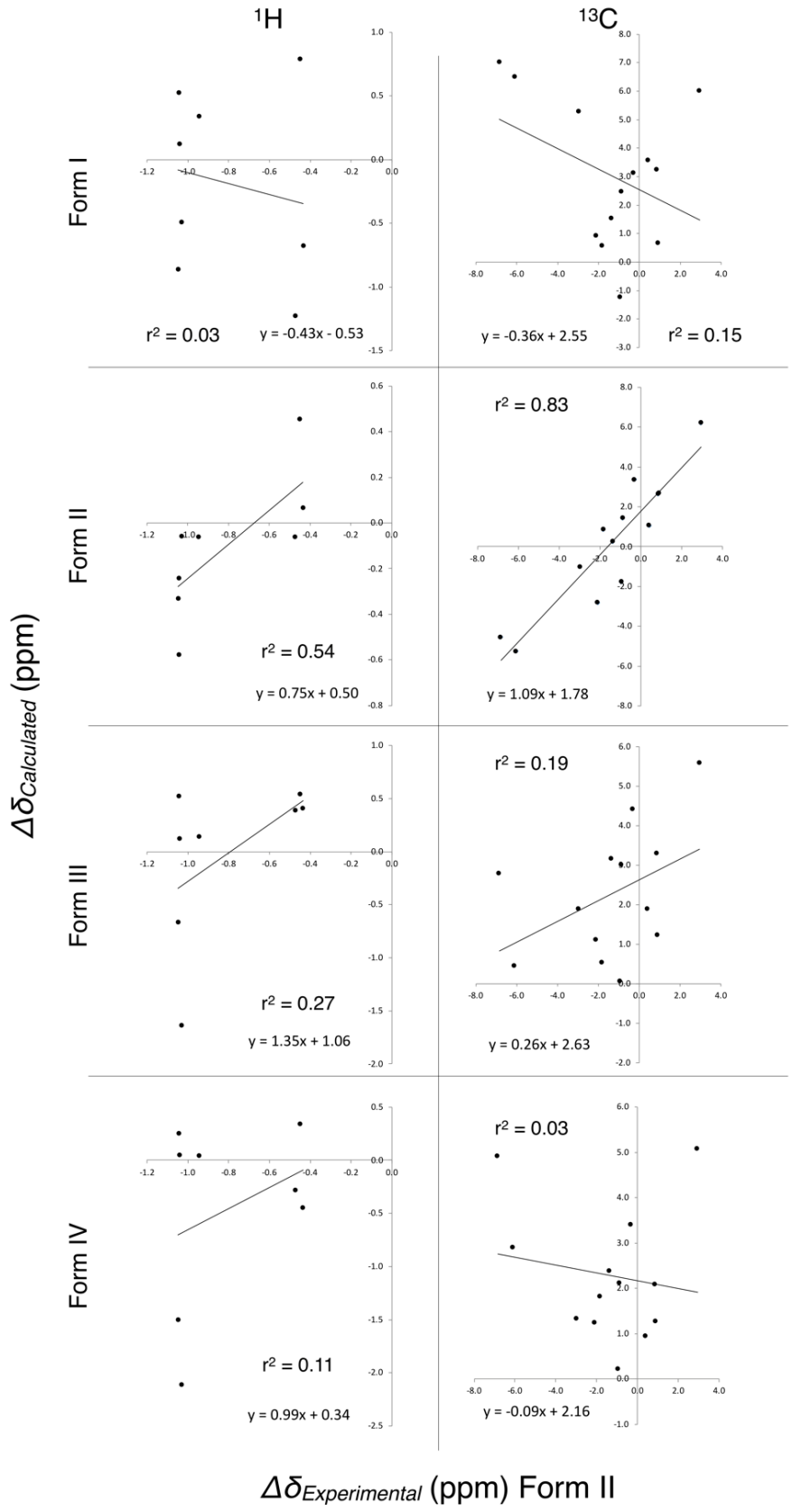


Figure 7. Graph of $\Delta\delta_{Calculated}$ for Forms I, II, III and IV against $\Delta\delta_{Experimental}$ for Form II for both ^{1}H and ^{13}C showing that the approach clearly discriminates Form II (r^2 = 0.54 and 0.83, respectively) from the other forms. See Tables 2 and 3 for data.

Supplementary information Figures S8 and S9 show the graphs of $\Delta\delta_{Calculated}$ for each conformation of Form III and Form IV against $\Delta\delta_{Experimental}$ for Form I and Form II, respectively; when both the ^{1}H and ^{13}C graphs are considered together, none would be identified as the correct form. This is of interest to how the approach is discriminating between forms because Form II has a similar conformation to those of Forms III and IV, yet none of their conformations correlate well against Form II experimental data, *i.e.* packing effects on the change in chemical shifts are important.

Table 3 provides a summary of all the fits (coefficient of determination r^2, gradient m, and intercept c values) from the regression analyses and corresponding p-values from the t-tests. This highlights that, under our approach, only Form I calculated data matches Form I experimental data, and only Form II calculated data matches Form II experimental data.

These calculations show that the approach allows ready discrimination of the correct crystal structure from several other very similar structures for TFA. The method therefore offers much promise as a novel general way of ranking trial structures in a CSP study. It is noted that ^{13}C seems to be the most discriminating nucleus.

3.6 Analysis I: Comparison of experimental and calculated solid-state NMR chemical shifts and RMSE values

As described in the Introduction, current NMR crystallography approaches for evaluating putative solid-state structures are based on comparing experimental and calculated solid-state NMR chemical shifts. Specifically, for organic molecules, Emsley and co-workers have established a threshold root-mean-squared error (RMSE) of 0.33 ppm for ^{1}H.[27]

Accordingly, Figures 8 and 9 plot the calculated ($\delta_{Solid\ calc}$) and experimentally measured ($\delta_{Solid\ expt}$) solid-state chemical shifts against each other, with Table 4 listing the linear regression analysis parameters. Table 4 also lists the RMSE values, showing that Form I and Form II are correctly identified below the ^{1}H threshold value (0.28 and 0.17 ppm), with a next highest RMSE value of 0.66 ppm. Table 4 also shows that the lowest ^{13}C RMSEs are also observed for the correct forms (1.34 and 1.28 ppm), but this is a less discriminatory measure (next highest RMSE value of 1.72 ppm).

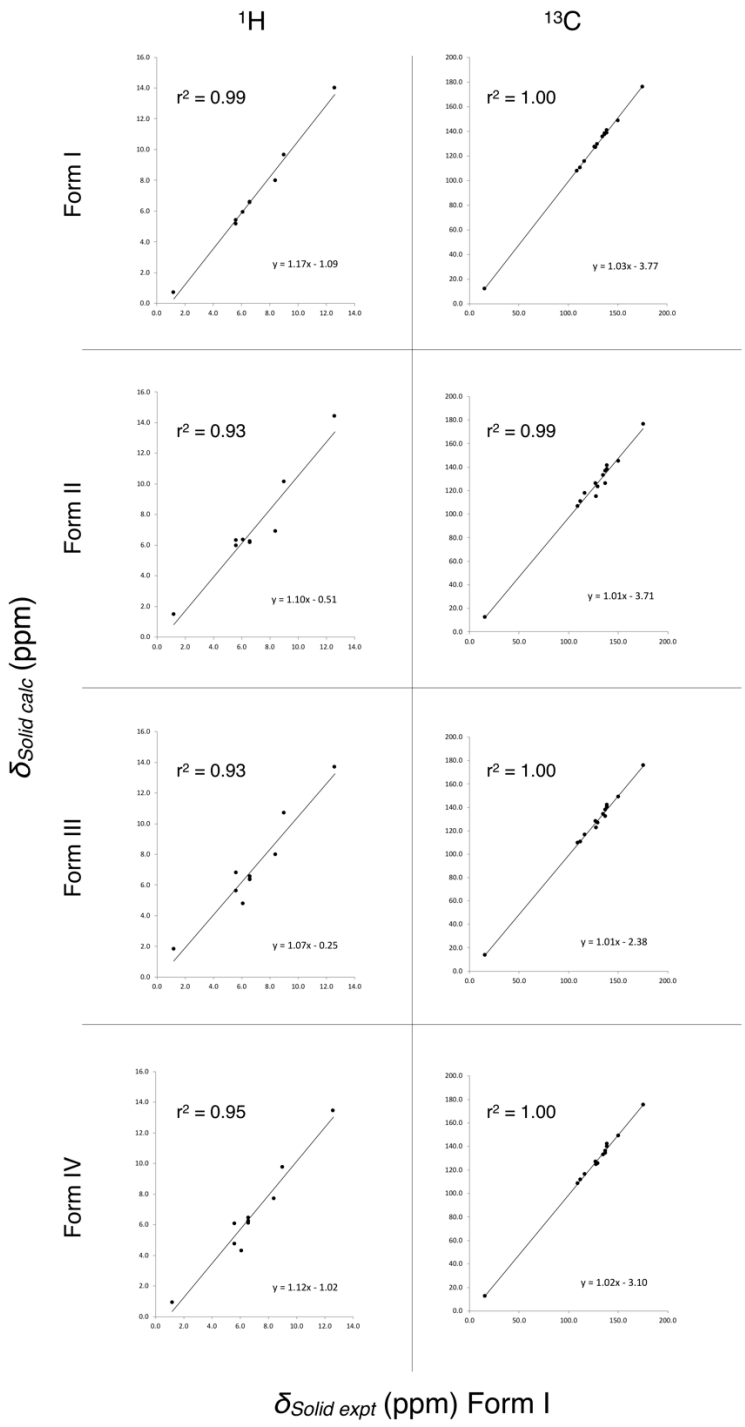


Figure 8 *– Graphs of $\delta_{Solid\ calc}$ for Forms I, II, III and IV against $\delta_{Solid\ expt}$ for Form I for both ^{1}H and ^{13}C. See Tables 1 and 4 for data.*

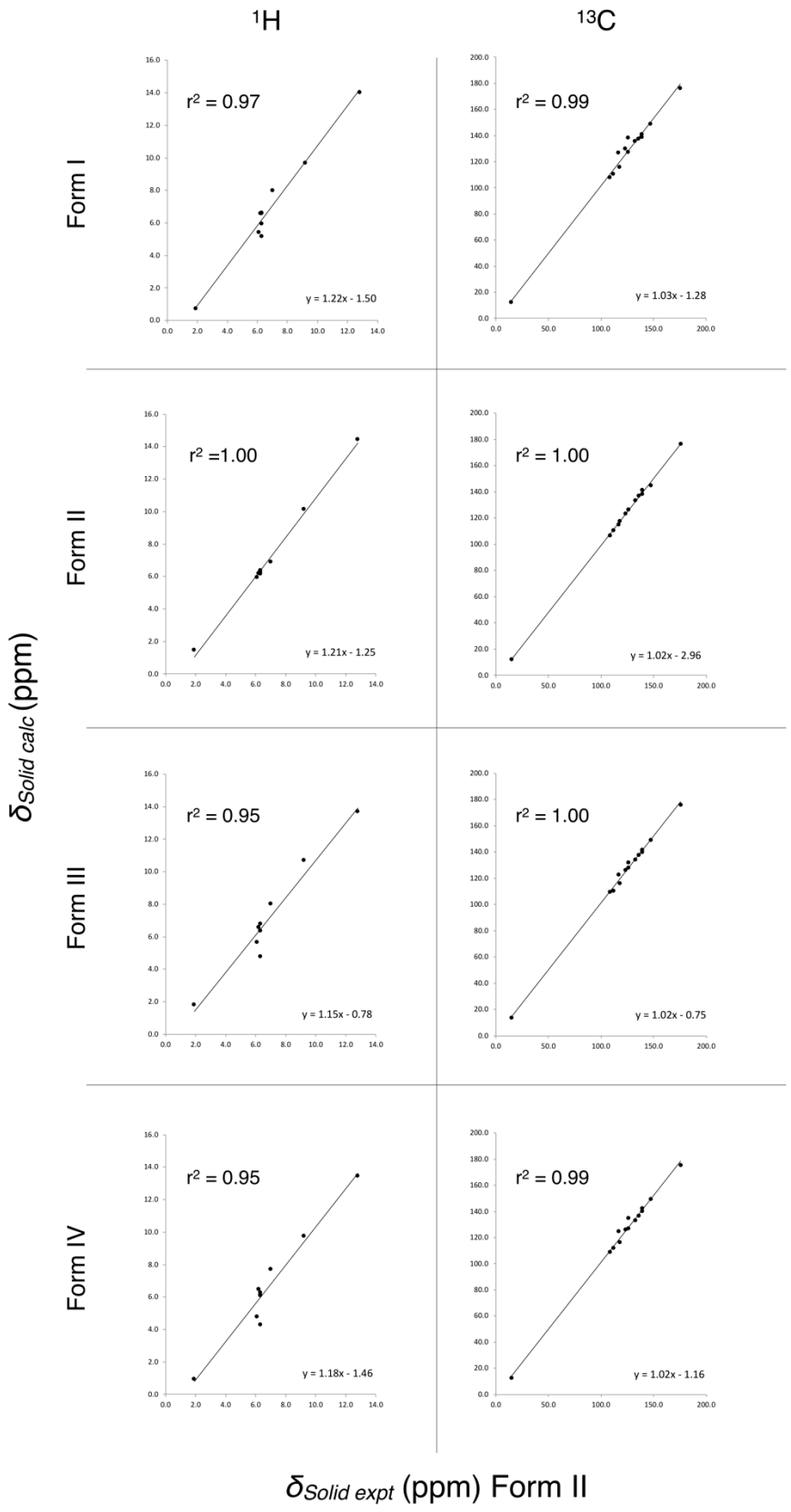


Figure 9 *– Graphs of $\delta_{Solid\ calc}$ for Forms I, II, III and IV against $\delta_{Solid\ expt}$ for Form II for both ^{1}H and ^{13}C. See Tables 1 and 4 for data.*

Table 4. Linear regression analysis parameters and RMSE values for comparing calculated (Form I, II, III and IV) and experimentally measured (Form I and II) chemical shifts in the solid state.

$\delta_{Solid\ calc}$ for:		$\delta_{Solid\ expt}$ for: Form I r^2	m	c	RMSE[a]	Form II r^2	m	c	RMSE[a]
^{1}H									
Form I		**0.99**[b]	**1.17**	**-1.09**	**0.28**	0.97	1.22	-1.50	0.75
Form II		0.93	1.10	-0.51	0.66	**1.00**	**1.21**	**-1.25**	**0.17**
Form III	mean	0.93	1.07	-0.25	0.69	0.95	1.15	-0.78	0.70
	1[c]	*0.93*	*1.10*	*-0.40*	*0.67*	*0.95*	*1.17*	*-0.92*	*0.74*
	2[d]	*0.91*	*1.05*	*-0.09*	*0.76*	*0.94*	*1.12*	*-0.64*	*0.71*
Form IV	mean	0.95	1.12	-1.02	0.79	0.95	1.18	-1.46	0.95
	1[e]	*0.96*	*1.10*	*-0.93*	*0.72*	*0.97*	*1.18*	*-1.46*	*0.74*
	2[f]	*0.93*	*1.10*	*-0.85*	*0.88*	*0.92*	*1.16*	*-1.25*	*1.07*
	3[g]	*0.94*	*1.14*	*-1.28*	*0.97*	*0.92*	*1.20*	*-1.67*	*1.19*
^{13}C									
Form I		**1.00**	**1.03**	**-3.77**	**1.34**	0.99	1.03	-1.28	4.94
Form II		0.99	1.01	-3.71	5.22	**1.00**	**1.02**	**-2.96**	**1.28**
Form III	mean	1.00	1.01	-2.38	2.33	1.00	1.02	-0.75	2.73
	1[c]	*0.99*	*1.02*	*-2.73*	*2.51*	*1.00*	*1.02*	*-1.22*	2.59
	2[d]	*1.00*	*1.01*	*-2.04*	*2.33*	*0.99*	*1.01*	*-0.28*	3.02
Form IV	mean	1.00	1.02	-3.10	1.82	0.99	1.02	-1.16	3.44
	1[e]	*1.00*	*1.02*	*-3.46*	*1.72*	*0.99*	*1.02*	*-1.33*	3.88
	2[f]	*1.00*	*1.02*	*-2.75*	*1.77*	*0.99*	*1.02*	*-0.78*	3.60
	3[g]	*1.00*	*1.02*	*-3.08*	*2.35*	*0.99*	*1.02*	*-1.35*	3.04

[a]RMSE values in ppm between $\delta_{Solid\ exp}$ and $\delta_{Solid\ calc}$ as calculated according to the approach of Emsley and co-workers.[27] p-values for the null hypothesis that m = 0, and the alternative hypothesis that m > 0, are <0.001 in all cases except for Form II $\delta_{Solid\ exp}$ vs with $\delta_{Solid\ calc}$ Form IV 2 ^{1}H (0.002), 3 1H (0.003) and Form I ^{13}C (0.002).The lower bound of the one-sided 95% confidence intervals for the correlation between $\Delta\delta_{Experimental}$ and $\Delta\delta_{Calculated}$ are given in supplementary Table S7.

[b]Values in bold indicate the fit parameters for the form corresponding to the measured experimental data, *i.e.* the ones the approach should identify (see Figures 8, 9, S10 and S11).

[c-g]R1 C7 – R1 N8 – R2 C1 – R2 C6 torsion angles are 126.8°, −138.4°, −115.8°, −126.0° and −134.2° in the diffraction structures, respectively.

3.7 Analysis II: Examining the choice of solvent and charge-state for solution NMR data

Table 3 shows the result of using the approach when taking the solution chemical shifts from $CDCl_3$. As noted above, the dynamic solution 3D structure was solved in water in the charged state and while these differences do not substantially change the conformational behaviour of TFA, they do substantially affect the chemical shifts relative to $CDCl_3$. Since the conformational behaviour of TFA is in fact not significantly affected by a wide range of organic solvents between these two extremes, we can now address the question of whether such differences in chemical shifts based on solvent and/or charge state have an impact on the discriminatory power of the approach.

Table 5 shows the coefficients of determination r^2 and p-values from graphs of $\Delta\delta_{Calculated}$ against $\Delta\delta_{Experimental}$ for a range of solvents and charge states on passing from solution to solid state for combinations of calculated (Form I, II, III and IV) and experimentally measured (Form I and II) changes in chemical shift (further solvents and full fit parameters are given in supplementary Table S8). In both water and methanol when TFA is in the charged state for both ^1H and ^{13}C, the r^2 and p-values clearly identify Form I, whilst Form II is only weakly identifiable. In contrast, in other solvents with TFA in the uncharged state (*i.e.* as in the crystal state), the approach always clearly identifies the correct forms; it therefore appears that, for TFA, matching the charge state is of high importance (especially for ^{13}C, as expected), whereas the choice of solvent is of surprisingly little importance. However, we note that solvents that mimic the packing environment such as benzene and toluene, which are structurally most similar to TFA, do nevertheless give the best fits, especially for ^1H. This suggests that detailed comparisons between the fit parameters for different solvents may allow fine packing-environment information to be inferred.

Another consideration is whether TFA is forming transiently-associated hydrogen-bonded dimers in apolar solvents such as $CDCl_3$. If this were the case, the solution chemical shifts ($\delta_{Solution\ expt}$) used in the approach would reflect these dimers, rather than the isolated monomers calculated for the approach employed in the paper. It is not possible to calculate chemical shifts reliably for solution dimers because their 3D-geometry is unknown in that they can be hydrogen-bonded either or both ways round, would likely be dynamically fluctuating across many conformations, and the populations of each conformation relative to the total are all unknown. Instead, by experimentally measuring chemical shifts at different concentrations and extrapolating to infinite dilution, it should be possible to determine chemical shifts representative of solution monomers, which would correspond better to the model used to calculate $\delta_{Solution\ calc}$. Chemical shift data for neutral TFA in $CDCl_3$ at 10 mM and 1 mM (Crick et al., manuscript in preparation) is available for all the ^{13}C atoms with directly attached ^1H. None of these chemical shifts move significantly with concentration ($^1H < 0.04$ ppm, $^{13}C < 0.2$ ppm), indicating that if transient dimer pairs are forming, they do not affect most of the molecule's chemical shifts. Data for neutral TFA in toluene-d_8 and benzene-d_6 at 10 mM and 1 mM, which includes the quaternary ^{13}C nuclei, likewise indicates that all changes are small ($^1H < 0.1$ ppm, $^{13}C < 0.3$ ppm) except for R1 C1 (the COOH carbon), which is 1.7 and 2.2 ppm lower at 1 mM compared to 10 mM, respectively. This indicates that transient hydrogen-bonded dimers are indeed likely forming in apolar solvents, but only manifest significant chemical shift effects at R1 C1. Therefore, presuming TFA in $CDCl_3$ behaves similarly to in toluene-d_8 and benzene-d_6, the $\delta_{Solution\ expt}$ ^{13}C chemical shifts for R1 C1 at infinite dilution would be expected to be approximately 2 ppm lower than the values used here in Table 1, giving $\Delta\delta_{Experimental}$ values in Table 2 for R1 C1 of 4.5 and 5.0 ppm (from 2.5 and 3.0 ppm), respectively, for Form I and Form II. These larger values would be in better agreement with the calculated values (6.0, 6.2 ppm). Since only the one ^{13}C chemical shift value is potentially affected by the presence of dimer pairs in $CDCl_3$, the analysis and conclusions from the approach above remains essentially unchanged.

Table 5. Coefficients of determination and p-values from graphs of $\Delta\delta_{Calculated}$ against $\Delta\delta_{Experimental}$ for TFA for calculations for Forms I-IV and experimental data for Forms I and II and in differing solvents.

$\Delta\delta_{Calculated}$ for:	r^2 correlation and p-value for $\Delta\delta_{Experimental}$ for solvent and Form against:																	
	CDCl3 TFA neutral				**Toluene-d8** TFA neutral				**CD_3OD, pH* 1.9**[a] TFA neutral				**CD_3OD, pH* 10.8** TFA charged				**D_2O, pH* 12.0** TFA charged			
	I		**II**		**I**		**II**		**I**		**II**		**I**		**II**		**I**		**II**	
	r^2	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**	**r^2**	**p**
^1H																				
Form I	**0.74**[b]	**0.003**	0.03	0.661	**0.85**	**0.001**	0.10	0.774	**0.75**	**0.003**	0.01	0.597	**0.85**	**<0.001**	0.00	0.463	**0.83**	**<0.001**	0.02	0.367
Form II	0.01	0.586	**0.54**	**0.019**	0.00	0.547	**0.73**	**0.004**	0.01	0.586	**0.53**	**0.020**	0.01	0.584	**0.74**	**0.003**	0.01	0.593	**0.60**	**0.012**
Form III (mean)	0.33	0.069	0.27	0.091	0.22	0.121	0.03	0.339	0.31	0.075	0.29	0.082	0.20	0.135	0.12	0.198	0.21	0.127	0.35	0.143
Form IV (mean)	0.55	**0.017**	0.11	0.211	0.46	**0.032**	0.00	0.495	0.53	**0.020**	0.13	0.191	0.42	**0.040**	0.05	0.292	0.42	0.040	0.16	0.222
^{13}C																				
Form I	**0.83**	**<0.001**	0.15	0.902	**0.87**	**<0.001**	0.11	0.868	**0.81**	**<0.001**	0.01	0.642	**0.42**	**0.008**	0.05	0.222	**0.49**	**0.004**	0.15	0.096
Form II	0.15	0.905	**0.83**	**<0.001**	0.12	0.872	**0.87**	**<0.001**	0.06	0.781	**0.81**	**<0.001**	0.13	0.885	**0.06**	0.209	0.12	0.873	**0.03**	0.283
Form III (mean)	0.03	0.301	0.19	0.069	0.04	0.249	0.23	**0.050**	0.05	0.226	0.27	**0.033**	0.00	0.575	0.01	0.357	0.00	0.519	0.02	0.315
Form IV (mean)	0.22	0.053	0.03	0.700	0.25	**0.040**	0.01	0.645	0.27	**0.033**	0.00	0.452	0.06	0.216	0.00	0.484	0.10	0.141	0.03	0.293

[a]pH* indicates the observed pH value, uncorrected for solvent composition.

[b]Values in bold indicate the fit parameters for the form corresponding to the measured experimental data, *i.e.*, the ones the approach should identify.

[c]p-values are for the null hypothesis that m = 0, and the alternative hypothesis that m > 0. Values in bold and underlined indicate a rejection of the null hypothesis at a significance level of 0.050, suggesting a positive correlation. The lower bound of the one-sided 95% confidence intervals for the correlation between $\Delta\delta_{Experimental}$ and $\Delta\delta_{Calculated}$ are given in supplementary Table S9.

3.8 Analysis III: Examining the relative contribution of packing and conformation

Several factors could be contributing to the discriminatory power of the approach which must have its origin in the differences in chemical shifts calculated for each solid form. Firstly, with regards to hydrogen bonding, all the forms exhibit the same acid dimer motif and no other hydrogen bonds; therefore this does not contribute to the calculated differences. Secondly, the conformations of Forms II, III and IV (Z' = 1, 2 and 3, respectively) at the torsion angles R1 C7 – R1 N8 – R2 C1 – R2 C6 are all quite similar and are found within the +/- 135° major mode in solution, whereas Form I (Z' = 1) is different and is found within the +/- 70° minor solution mode (see supplementary Figures S12 and S13, and Table S10); there is very little difference at the R1 O1A – R1 C1 – R1 C2 – R1 C7 and R1 C2 – R1 C7 – R1 N8 – R2 C1 torsions. Thirdly, the crystal packing is different between the Forms (see supplementary Table S11). Using the crystal packing similarity tool within Mercury[37] all forms were compared with each other. Forms III and IV are similar to each other having 11 of 15 molecules in common locations. However, Forms III and IV are quite different to both Form I (2 of 15) and Form II (2 of 15), and furthermore Form I and Form II also have quite different packing (2 of 15).

In an attempt to assess the relative contribution of conformational and packing effects on chemical shifts to the success of the approach, we calculated isolated molecule chemical shifts for each form in an enlarged unit cell $\delta_{Isolated\ calc}$, *i.e.* removing all inter-molecular contacts (see supplementary Table S12).[52-56] Replacing $\delta_{Solid\ calc}$ in equation 2 with this $\delta_{Isolated\ calc}$ allows us to calculate a modified $\Delta\delta_{Calculated}$ value, $\Delta\delta_{Calculated\text{-}isolated}$, to plot against the experimentally observed changes. Supplementary Figures S14 and S15 show $\Delta\delta_{Calculated\text{-}Isolated}$ for Forms I, II, III and IV against $\Delta\delta_{Experimental}$ for Form I and II, respectively for both ^{1}H and ^{13}C (again with R1 HN and R2 C5 omitted).

In Table 6, we provide a summary of all the fits' parameters (slope m, intercept c and coefficient of determination r^2) and their respective p-values from the modified approach for these isolated crystal conformations. In the case of Form I, the ^{13}C correlations (r^2 = 0.46, p = 0.006) still successfully discriminate Form I from the other forms (r^2 = 0.30, 0.07 and 0.11; p = 0.973, 0.812, 0.025), whereas ^{1}H is now poorly discriminating (r^2 = positive correlation of 0.03 versus negative correlations of 0.01, 0.17 and 0.32, p = 0.334, 0.611, 0.841 and 0.631). Similarly, for Form II, ^{13}C is still discriminating (r^2 = 0.78, p < 0.001), while ^{1}H is not. Indeed, and perhaps surprisingly, the ^{13}C still discriminates Form II from Forms III (r^2 = 0.30, p = 0.026) and IV (r^2 = 0.50, p = 0.997) despite them all having similar conformations in the major mode (see supplementary Figure S13, comparing arrow positions); Form III has the more similar conformations to Form II and this is reflected in the more positive correlation and significant p-value in Table 6.

This indicates that both conformation and packing effects contribute to the discriminatory power of the approach with, at least for TFA, seemingly ^{1}H being more sensitive to packing, and ^{13}C more sensitive to conformation. We consider it remarkable that it is still possible to identify the correct crystal structure when the effect of packing on chemical shift is removed. Moreover, this implies that, for TFA, it may be possible to predict the crystal conformation from ^{13}C chemical shift data without any prior knowledge of the crystal structure.

Table 6. Linear regression analysis parameters and p-values from graphs of $\Delta\delta_{Calculated\text{-}Isolated}$ against $\Delta\delta_{Experimental}$ (i.e., with $\delta_{Solid\ calc}$ replaced with $\delta_{Isolated\ calc}$, a calculation for an isolated molecule extracted from the crystal structure, see Figures S14 and S15) on passing from solution to solid state for combinations of calculated (Form I, II, III and IV) and experimentally measured (Form I and II) changes in chemical shift for isolated crystal conformations.

		$\Delta\delta_{Experimental}$ for:							
		Form I				Form II			
$\Delta\delta_{Calculated\text{-}Isolated}$ for:		r^2	m	c	p-value[b]	r^2	m	c	p-value[b]
^{1}H									
Form I		**0.03**[a]	**0.11**	**-0.20**	0.334	0.23	-0.67	-0.83	0.888
Form II		0.01	-0.03	-0.05	0.611	**0.26**	**0.27**	**0.19**	0.098
Form III	mean	0.17	-0.05	-0.06	0.841	0.41	0.17	0.11	**0.042**
	1[c]	*0.40*	*-0.09*	*-0.09*	*0.953*	*0.37*	*0.19*	*0.13*	*0.056*
	2[d]	*0.00*	*-0.01*	*-0.03*	*0.553*	*0.33*	*0.14*	*0.10*	*0.066*
Form IV	mean	0.32	-0.04	-0.10	0.631	0.10	-0.06	-0.13	0.772
	1[e]	*0.40*	*-0.09*	*-0.09*	*0.623*	*0.30*	*-0.13*	*-0.16*	*0.920*
	2[f]	*0.06*	*-0.03*	*-0.13*	*0.719*	*0.11*	*-0.10*	*-0.18*	*0.790*
	3[g]	*0.01*	*-0.08*	*0.02*	*0.381*	*0.06*	*0.04*	*-0.06*	*0.274*
^{13}C									
Form I		**0.46**	**0.76**	**0.05**	**0.006**	0.64	-0.67	0.02	1.000
Form II		0.30	-0.42	-0.73	0.973	**0.78**	**0.51**	**-0.51**	**<0.001**
Form III	mean	0.07	-0.08	-0.27	0.812	0.30	0.12	-0.19	**0.026**
	1[c]	*0.01*	*0.03*	*-0.58*	*0.371*	*0.27*	*0.15*	*-0.32*	***0.033***
	2[d]	*0.24*	*-0.19*	*0.05*	*0.954*	*0.09*	*0.09*	*-0.06*	*0.167*
Form IV	mean	0.11	0.11	0.15	**0.025**	0.50	-0.287	-0.13	0.997
	1[e]	*0.01*	*0.04*	*-0.58*	***<0.001***	*0.42*	*-0.36*	*-0.21*	*0.992*
	2[f]	*0.41*	*0.40*	*-0.18*	***0.009***	*0.55*	*-0.34*	*-0.19*	*0.998*
	3[g]	*0.05*	*-0.11*	*0.32*	*0.759*	*0.10*	*-0.12*	*0.02*	*0.855*

[a]Values in bold indicate the fit parameters for the Form corresponding to the measured experimental data, *i.e.* the ones the approach should identify (see supplementary Figures S14 and S15).

[b]p-values are for the null hypothesis that m = 0, and the alternative hypothesis that m > 0. Values in bold and underlined indicate a rejection of the null hypothesis at a significance level of 0.050, suggesting a positive correlation. The lower bound of the one-sided 95% confidence intervals for the correlation between $\Delta\delta_{Experimental}$ and $\Delta\delta_{Calculated}$ are given in supplementary Table S13.

[c-g]R1 C7 – R1 N8 – R2 C1 – R2 C6 torsion angles are 126.8°, −138.4°, −115.8°, −126.0° and −134.2° in the diffraction structures, respectively.

3.9 Analysis IV: Using the approach when there is no crystal structure

To explore this hypothesis, we used the calculated ^{1}H and ^{13}C chemical shifts for isolated conformations with the torsion angle R1 C7 – R1 N8 – R2 C1 – R2 C6 varied at 15° intervals and calculated their changes in chemical shift, giving a set of $\Delta\delta_{Calculated\text{-}Isolated\ conformation\ at\ specific\ 15°\ angle}$ values (*i.e*. from Supplementary Table S4D, replacing $\delta_{Solid\ calc}$ by $\delta_{Isolated\ conformation\ at\ specific\ 15°\ angle}$ for each 15° angle in equation 2). The respective fit parameters (slope m, intercept c and coefficient of determination r^2 values) of each of these $\Delta\delta_{Calculated\text{-}Isolated\ conformation\ at\ specific\ 15°\ angle}$ values against Form I and Form II experimental ^{1}H and ^{13}C chemical shift data ($\Delta\delta_{Experimental}$ from Table 2) were then determined as before. The variations in m and r^2 with torsion angle R1 C7 – R1 N8 – R2 C1 – R2 C6 are shown in supplementary Figure S16, and clearly differ between Forms I and II.

Since higher r^2 values with positive correlations (*i.e.* m > 1) indicate reasonable fits of a given conformation to the experimental data, a single metric for how well each trial conformation matches the experimental data (and its underlying conformation) can be produced by using the correlation coefficient r (which includes both the strength of the correlation and its sign) as shown in Figure 10. The highest points on these graphs therefore represent the conformations that are most consistent with the experimental chemical shift data for each polymorph. Figure 10 shows that, for ^{13}C chemical shift $\Delta\delta_{Experimental}$ data, the range of torsion angles most consistent with the experimental data are ±75 to ±120° and ±135 to ±180° for Form I and II, respectively. Significantly, the experimentally observed conformations for each form are indeed within these regions (Form I = ±74.9° and Form II = ±142.6°). The ^{1}H chemical shift $\Delta\delta_{Experimental}$ data indicate similar ranges (Form I ±30 to ±90°, Form II ±135 to ±165°), albeit with less certainty, consistent with the observation above that the ^{1}H data is less sensitive to conformation (supplementary Figure S17).

These results demonstrate that the conformation present in each TFA form can be considerably narrowed to a specific range solely using solution- and solid-state NMR data, without any prior knowledge of the crystal structure.

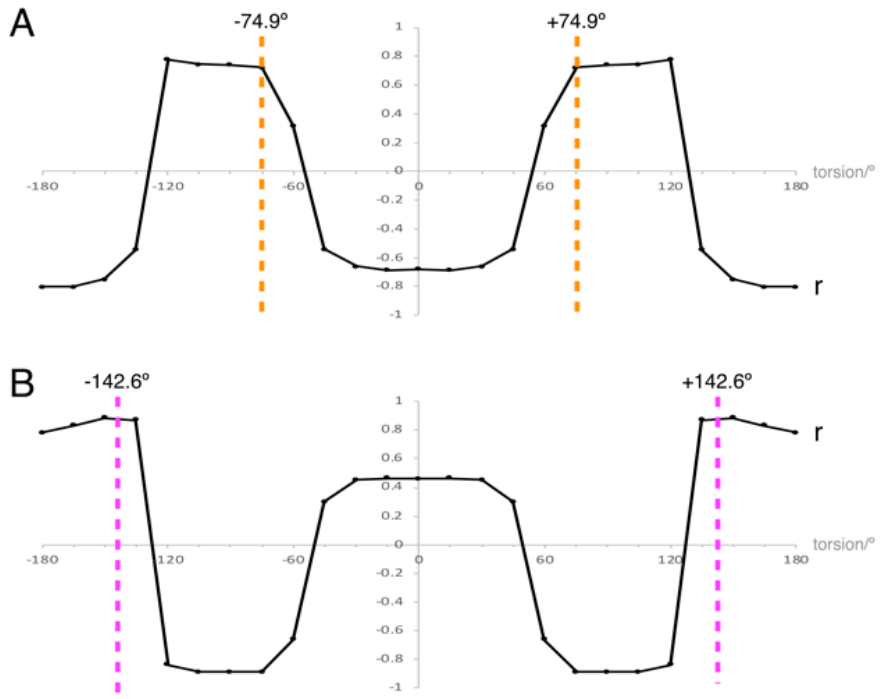


Figure 10. Correlation coefficients r for all $\Delta\delta_{Calculated\text{-}Isolated\ conformation\ at\ specific\ 15°\ angle}$ values against A) Form I and B) Form II experimental ^{13}C chemical shift $\Delta\delta_{Experimental}$ data. Conformations differ by 15° intervals at torsion angle R1 C7 – R1 N8 – R2 C1 – R2 C6. Form I and Form II conformations are indicated with dashed orange and magenta lines, respectively.

4 Conclusions

We have proposed a novel approach for quantitatively assessing putative crystal structures using a combination of experimental solution- and solid-state NMR data and DFT based GIPAW calculations, and have applied it to TFA as a proof-of-concept exemplar. The approach provides a quantitative method for assessing how well a suggested structure matches the experimental solid-state NMR chemical shifts. For TFA, our approach can discriminate between four polymorphs of which Forms II, III and IV are structurally similar, with ^{13}C more discriminating than ^{1}H. Moreover, we have demonstrated that, even without prior knowledge of the crystal structure, we can use our approach to predict a narrow range for the conformation present for each solid form of TFA. We recognise that this approach will only work for torsions whose chemical shifts are sensitive to changes in conformation and that in other systems some torsions may be insensitive.

The established approach of comparing GIPAW calculated and experimental solid-state chemical shifts alone via RMSE values[27] also identifies the correct form but, in contrast to our approach, with greater discriminating power for ^{1}H as compared to ^{13}C. This difference shows complementarity of the two approaches that, when applied in combination, offers further reassurance of a correct match. In conclusion, we see potential to integrate our new approach into workflows for the many cases of solid-state chemistry where crystal structures are not available. Consider the application of CSP to solve intractable crystal structures. Specifically, before starting a CSP campaign, the conformational search space could be significantly reduced by using the experimental data to predict the most likely conformations present. After the CSP campaign has returned putative crystal structures, the method could be used for a second time with the experimental data in this iteration permitting reliable and quantified identification of correct, plausible and incorrect structures. An extension of this proof-of-concept work could be to apply the approach also to a consideration of low-energy structures resulting from a CSP campaign, as for example available for TFA from Ref.[36] We can also envisage that this approach could be used in combination with structure determination by PXRD.

Supporting Information. PXRD analysis; comparison of crystal structures; conformational behaviour of TFA in solution, and comparison to torsion angles in the solid-state forms; GIPAW calculated chemical shifts TFA in an isolated box with varied torsion angle; calculation of $\delta_{\text{Solution calc}}$; MAS NMR spectra; regression analysis plots of $\Delta\delta_{Calculated}$ for each conformation of Form III and Form IV Form II against $\Delta\delta_{Experimental}$ for forms I and II; plots of $\delta_{Solid\ calc}$ against $\delta_{Solid\ expt}$; lower bounds of 95% confidence intervals for correlations; regression analysis parameters for a range of solvents; crystal packing similarity; plots of $\Delta\delta_{Calculated\text{-}Isolated}$ for Forms I and II against $\Delta\delta_{Experimental}$ for Forms I, II, III and IV; plots of regression analysis parameters for $\Delta\delta_{Calculated\text{-}Isolated\ conformation\ at\ specific\ 15°\ angle}$ against torsion angle; schema of example workflows for using the approach in CSP; atom numbering in CASTEP output files (pdf)

Acknowledgements

AKM thanks the University of Warwick for a Chancellor's International Scholarship. The experimental (NMR and PXRD) and calculated (.magres, see supplementary Figure S18) data for this study are provided at WRAP, the Warwick Research Archive Portal http://wrap.warwick.ac.uk/***

References

(1) Harris, K. D. M.; Johnston, R. L.; Kariuki, B. M.; Tedesco, E.; Turner, G. W. In *Epdic 7: European Powder Diffraction, Pts 1 And 2*; Trans Tech Publications Ltd: Zurich-Uetikon, 2001; Vol. 378-3, pp 38-46.

(2) Tedesco, E.; Dhillon, S. S.; Harris, K. D. M.; Johnston, R. L.; Turner, G. W.; Kariuki, B. M. In *Epdic 7: European Powder Diffraction, Pts 1 And 2*; Trans Tech Publications Ltd: Zurich-Uetikon, 2001; Vol. 378-3, pp 784-788.

(3) Brunelli, M.; Wright, J. P.; Vaughan, G. R. M.; Mora, A. J.; Fitch, A. N. Solving larger molecular crystal structures from powder diffraction data by exploiting anisotropic thermal expansion *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 2029-2032.

(4) Smrcok, L.; Brunelli, M.; Boca, M.; Kucharik, M. Structure of K_2TaF_7 at 993 K: the combined use of synchrotron powder data and solid-state DFT calculations *J. Appl. Crystallogr.* **2008**, *41*, 634-636.

(5) Harris, R. K. NMR crystallography: the use of chemical shifts *Solid State Sci.* **2004**, *6*, 1025-1037.

(6) Elena, B.; Pintacuda, G.; Mifsud, N.; Emsley, L. Molecular structure determination in powders by NMR crystallography from proton spin diffusion *J. Am. Chem. Soc.* **2006**, *128*, 9555-9560.

(7) Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V. Chemical shift computations on a crystallographic basis: some reflections and comments *Magn. Reson. Chem.* **2007**, *45*, S174-S186.

(8) Taulelle, F. In *Encyclopedia of Magnetic Resonance*; Harris, R. K., Wasylishen, R., Eds.; John Wiley: Chichester, 2009; p DOI:10.1002/9780470034590.emrstm9780470031003.

(9) Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. Powder NMR crystallography of thymol *Phys. Chem. Chem. Phys.* **2009**, *11*, 2610-2621.

(10) Charpentier, T. The PAW/GIPAW approach for computing NMR parameters: A new dimension added to NMR study of solids *Solid State Nucl. Magn. Reson.* **2011**, *40*, 1-20.

(11) Bonhomme, C.; Gervais, C.; Babonneau, F.; Coelho, C.; Pourpoint, F.; Azais, T.; Ashbrook, S. E.; Griffin, J. M.; Yates, J. R.; Mauri, F. et al First-Principles Calculation of NMR Parameters Using the Gauge Including Projector Augmented Wave Method: A Chemist's Point of View *Chem. Rev.* **2012**, *112*, 5733-5779.

(12) Ashbrook, S. E.; McKay, D. Combining solid-state NMR spectroscopy with first-principles calculations - a guide to NMR crystallography *Chem. Commun.* **2016**, *52*, 7186-7204.

(13) Bryce, D. L. NMR crystallography: structure and properties of materials from solid-state nuclear magnetic resonance observables *IUCrJ* **2017**, *4*, 350-359.

(14) Hughes, C. E.; Reddy, G. N. M.; Masiero, S.; Brown, S. P.; Williams, P. A.; Harris, K. D. M. Determination of a complex crystal structure in the absence of single crystals: analysis of powder X-ray diffraction data, guided by solid-state NMR and periodic DFT calculations, reveals a new 2 '-deoxyguanosine structural motif *Chem. Sci.* **2017**, *8*, 3971-3979.

(15) Tatton, A. S.; Blade, H.; Brown, S. P.; Hodgkinson, P.; Hughes, L. P.; Lill, S. O. N.; Yates, J. R. Improving Confidence in Crystal Structure Solutions Using NMR Crystallography: The Case of β-Piroxicam *Cryst. Growth Des.* **2018**, *18*, 3339-3351.

(16) Day, G. M.; Motherwell, W. D. S.; Jones, W. A strategy for predicting the crystal structures of flexible molecules: the polymorphism of phenobarbital *Phys. Chem. Chem. Phys.* **2007**, *9*, 1693-1704.

(17) Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. A major advance in crystal structure prediction *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 2427-2430.

(18) Neumann, M. A. Tailor-made force fields for crystal-structure prediction *J. Phys. Chem. B* **2008**, *112*, 9810-9829.

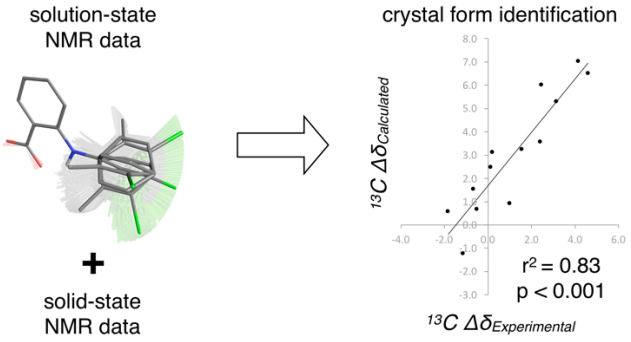
(19) Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; van de Streek, J. Progress in Crystal Structure Prediction *Chem.-Eur. J.* **2011**, *17*, 10736-10744.

(20) van de Streek, J.; Neumann, M. A. Crystal-structure prediction of pyridine with four independent molecules *Crystengcomm* **2011**, *13*, 7135-7142.
(21) Day, G. M. Current approaches to predicting molecular organic crystal structures *Crystallogr. Rev* **2011**, *17*, 3-52.
(22) Price, S. L. Why don't we find more polymorphs? *Acta Crystallogr. Sect. B-Struct. Sci.Cryst. Eng. Mat.* **2013**, *69*, 313-328.
(23) Price, S. L. Predicting crystal structures of organic compounds *Chem. Soc. Rev.* **2014**, *43*, 2098-2111.
(24) Neumann, M. A.; de Streek, J. V.; Fabbiani, F. P. A.; Hidber, P.; Grassmann, O. Combined crystal structure prediction and high-pressure crystallization in rational pharmaceutical polymorph screening *Nat. Commun.* **2015**, *6*, 7793.
(25) Price, S. L.; Braun, D. E.; Reutzel-Edens, S. M. Can computed crystal energy landscapes help understand pharmaceutical solids? *Chem. Commun.* **2016**, *52*, 7065-7077.
(26) Pickard, C. J.; Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts *Phys. Rev. B* **2001**, *63*, 245101.
(27) Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. Powder Crystallography by Combined Crystal Structure Prediction and High-Resolution ^{1}H Solid-State NMR Spectroscopy *J. Am. Chem. Soc.* **2010**, *132*, 2564-2566.
(28) Baias, M.; Widdifield, C. M.; Dumez, J. N.; Thompson, H. P. G.; Cooper, T. G.; Salager, E.; Bassil, S.; Stein, R. S.; Lesage, A.; Day, G. M. et al Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state ^{1}H NMR spectroscopy *Phys. Chem. Chem. Phys.* **2013**, *15*, 8069-8080.
(29) Hofstetter, A.; Balodis, M.; Paruzzo, F. M.; Widdifield, C. M.; Steyanato, G.; Pinon, A. C.; Bygrave, P. J.; Day, G. M.; Emsley, L. Rapid Structure Determination of Molecular Solids Using Chemical Shifts Directed by Unambiguous Prior Constraints *J. Am. Chem. Soc.* **2019**, *141*, 16624-16634.
(30) Santos, S. M.; Rocha, J.; Mafra, L. NMR Crystallography: Toward Chemical Shift-Driven Crystal Structure Determination of the beta-Lactam Antibiotic Amoxicillin Trihydrate *Cryst. Growth Des.* **2013**, *13*, 2390-2395.
(31) Baias, M.; Dumez, J. N.; Svensson, P. H.; Schantz, S.; Day, G. M.; Emsley, L. De Novo Determination of the Crystal Structure of a Large Drug Molecule by Crystal Structure Prediction-Based Powder NMR Crystallography *J. Am. Chem. Soc.* **2013**, *135*, 17501-17507.
(32) Paruzzo, F. M.; Hofstetter, A.; Musil, F.; De, S.; Ceriotti, M.; Emsley, L. Chemical shifts in molecular solids by machine learning *Nat. Commun.* **2018**, *9*, 4501.
(33) Cruz-Cabeza, A. J.; Bernstein, J. Conformational Polymorphism *Chem. Rev.* **2014**, *114*, 2170-2191.
(34) Lopez-Mejias, V.; Kampf, J. W.; Matzger, A. J. Polymer-Induced Heteronucleation of Tolfenamic Acid: Structural Investigation of a Pentamorph *J. Am. Chem. Soc.* **2009**, *131*, 4554-4555.
(35) Andersen, K. V.; Larsen, S.; Alhede, B.; Gelting, N.; Buchardt, O. Characterization of 2 polymorphic forms of tolfenamic acid, N-(2-methyl-3-chlorophenyl)anthranilic acid - their crystal-structures and relative stabilities *J. Chem. Soc.-Perkin Trans. 2* **1989**, 1443-1447.
(36) Case, D. H.; Srirambhatla, V. K.; Guo, R.; Watson, R. E.; Price, L. S.; Polyzois, H.; Cockcroft, J. K.; Florence, A. J.; Tocher, D. A.; Price, S. L. Successful Computationally Directed Templating of Metastable Pharmaceutical Polymorphs *Cryst. Growth Des.* **2018**, *18*, 5322-5331.
(37) Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; van De Streek, J. Mercury: visualization and analysis of crystal structures *J. Appl. Crystallogr.* **2006**, *39*, 453-457.
(38) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database *Acta Crystallogr. Sect. B-Struct. Sci.Cryst. Eng. Mat.* **2016**, *72*, 171-179.

(39) Du, W.; Cruz-Cabeza, A. J.; Woutersen, S.; Davey, R. J.; Yin, Q. Can the study of self-assembly in solution lead to a good model for the nucleation pathway? The case of tolfenamic acid *Chem. Sci.* **2015**, *6*, 3515-3524.
(40) Metz, G.; Wu, X. L.; Smith, S. O. Ramped-Amplitude Cross-Polarization In Magic-Angle-Spinning NMR *J. Magn. Reson. Ser. A* **1994**, *110*, 219-227.
(41) Fung, B. M.; Khitrin, A. K.; Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids *J. Magn. Reson.* **2000**, *142*, 97-101.
(42) Elena, B.; Lesage, A.; Steuernagel, S.; Bockmann, A.; Emsley, L. Proton to carbon-13 INEPT in solid-state NMR spectroscopy *J. Am. Chem. Soc.* **2005**, *127*, 17296-17302.
(43) Elena, B.; de Paepe, G.; Emsley, L. Direct spectral optimisation of proton-proton homonuclear dipolar decoupling in solid-state NMR *Chem. Phys. Lett.* **2004**, *398*, 532-538.
(44) Hayashi, S.; Hayamizu, K. Chemical-shift standards in high-resolution solid-state NMR (1) ^{13}C, ^{29}Si and ^{1}H nuclei *Bull. Chem. Soc. Jpn.* **1991**, *64*, 685-687.
(45) Morcombe, C. R.; Zilm, K. W. Chemical shift referencing in MAS solid state NMR *J. Magn. Reson.* **2003**, *162*, 479-486.
(46) Harris, R. K.; Becker, E. D.; De Menezes, S. M. C.; Granger, P.; Hoffman, R. E.; Zilm, K. W. Further conventions for NMR shielding and chemical shifts (IUPAC recommendations 2008) *Pure Appl. Chem.* **2008**, *80*, 59-84.
(47) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP *Z. Kristall.* **2005**, *220*, 567-570.
(48) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
(49) Reddy, G. N. M.; Cook, D. S.; Iuga, D.; Walton, R. I.; Marsh, A.; Brown, S. P. An NMR crystallography study of the hemihydrate of 2 ', 3 '-O-isopropylidineguanosine *Solid State Nucl. Magn. Reson.* **2015**, *65*, 41-48.
(50) Hartman, J. D.; Kudla, R. A.; Day, G. M.; Mueller, L. J.; Beran, G. J. O. Benchmark fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals *Phys. Chem. Chem. Phys.* **2016**, *18*, 21686-21709.
(51) Spek, A. L. Structure validation in chemical crystallography *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2009**, *65*, 148-155.
(52) Gervais, C.; Profeta, M.; Lafond, V.; Bonhomme, C.; Azais, T.; Mutin, H.; Pickard, C. J.; Mauri, F.; Babonneau, F. Combined ab initio computational and experimental multinuclear solid-state magnetic resonance study of phenylphosphonic acid *Magn. Reson. Chem.* **2004**, *42*, 445-452.
(53) Yates, J. R.; Pham, T. N.; Pickard, C. J.; Mauri, F.; Amado, A. M.; Gil, A. M.; Brown, S. P. An investigation of weak CH...O hydrogen bonds in maltose anomers by a combination of calculation and experimental solid-state NMR spectroscopy *J. Am. Chem. Soc.* **2005**, *127*, 10216-10220.
(54) Schmidt, J.; Hoffmann, A.; Spiess, H. W.; Sebastiani, D. Bulk chemical shifts in hydrogen-bonded systems from first-principles calculations and solid-state-NMR *J. Phys. Chem. B* **2006**, *110*, 23204-23210.
(55) Mafra, L.; Santos, S. M.; Siegel, R.; Alves, I.; Paz, F. A. A.; Dudenko, D.; Spiess, H. W. Packing Interactions in Hydrated and Anhydrous Forms of the Antibiotic Ciprofloxacin: a Solid-State NMR, X-ray Diffraction, and Computer Simulation Study *J. Am. Chem. Soc.* **2012**, *134*, 71-74.
(56) Zilka, M.; Sturniolo, S.; Brown, S. P.; Yates, J. R. Visualising crystal packing interactions in solid-state NMR: Concepts and applications *J. Chem. Phys.* **2017**, *147*.
(57) Sturniolo, S.; Green, T. F. G.; Hanson, R. M.; Zilka, M.; Refson, K.; Hodgkinson, P.; Brown, S. P.; Yates, J. R. Visualization and processing of computed solid-state NMR parameters: MagresView and MagresPython *Solid State Nucl. Magn. Reson.* **2016**, *78*, 64-70.

(58) Blundell, C. D.; Packer, M. J.; Almond, A. Quantification of free ligand conformational preferences by NMR and their relationship to the bioactive conformation *Bioorg. Med. Chem.* **2013**, *21*, 4976-4987.
(59) R Foundation for Statistical Computing: Vienna, Austria, 2020.

TOC Graphic


solution-state
NMR data
+
solid-state
NMR data
crystal form identification
$^{13}C\ \Delta\delta_{Calculated}$
$^{13}C\ \Delta\delta_{Experimental}$
$r^2 = 0.83$
$p < 0.001$